HISTORY,

NUMBERS,

and

WAR

A HERO journal

Volume 1 Number 1 **Spring 1977**

T N Dupuy Associates

THE LANCHESTER EQUATIONS AND HISTORICAL WARFARE: AN ANALYSIS OF SIXTY WORLD WAR II LAND ENGAGEMENTS

JANICE B. FAIN

Background and Objectives

The method by which combat losses are computed is one of the most critical parts of any combat model. The Lanchester equations, which state that a unit's combat losses depend on the size of its opponent, are widely used for this purpose.

In addition to their use in complex dynamic simulations of warfare, the Lanchester equations have also served as simple mathematical models. In fact, during the last decade or so there has been an explosion of theoretical developments based on them. By now their variations and modifications are numerous, and "Lanchester theory" has become almost a separate branch of applied mathematics. However, compared with the effort devoted to theoretical developments, there has been relatively little empirical testing of the basic thesis that combat losses are related to force sizes.

One of the first empirical studies of the Lanchester equations was Engel's classic work on the Iwo Jima campaign in which he found a reasonable fit between computed and actual U.S. casualties (Note 1). Later studies were somewhat less supportive (Notes 2 and 3), but an investigation of Korean war battles showed that, when the simulated combat units were constrained to follow the tactics of their historical counterparts, casualties during combat could be predicted to within 1 to 13 percent (Note 4).

Janice Bloom Fain, a Senior Associate of CACI, Inc., is a physicist whose special interests are in the applications of computer simulation techniques to industrial and military operations; she is the author of numerous reports and articles in this field. This paper was presented by Dr. Fain at the Military Operations Research Symposium at Fort Eustis, Virginia.

Taken together, these various studies suggest that, while the Lanchester equations may be poor descriptors of large battles extending over periods during which the forces were not constantly in combat, they may be adequate for predicting losses while the forces are actually engaged in fighting. The purpose of the work reported here is to investigate 60 carefully selected World War II engagements. Since the durations of these battles were short (typically two to three days), it was expected that the Lanchester equations would show a closer fit than was found in studies of larger battles. In particular, one of the objectives was to repeat, in part, Willard's work on battles of the historical past (Note 3).

The Data Base

Probably the most nearly complete and accurate collection of combat data is the data on World War II compiled by the Historical Evaluation and Research Organization (HERO). From their data HERO analysts selected, for quantitative analysis, the following 60 engagements from four major Italian campaigns:

- Salerno, 9-18 Sep 1943, 9 engagements
- Volturno, 12 Oct-8 Dec 1943, 20 engagements
- Anzio, 22 Jan-29 Feb 1944, 11 engagements
- Rome, 14 May-4 June 1944, 20 engagements

The complete data base is described in a HERO report (Note 5). The work described here is not the first analysis of these data. Statistical analyses of weapon effectiveness and the testing of a combat model (the Quantified Judgment Method, QJM) have been carried out (Note 6). The work discussed here examines these engagements from the viewpoint of the Lanchester equations to consider the question: "Are casualties during combat related to the numbers of men in the opposing forces?"

The variables chosen for this analysis are shown in Table 1. The "winners" of the engagements were specified by HERO on the basis of casualties suffered, distance advanced, and subjective estimates of the percentage of the commander's objective achieved. Variable 12, the Combat Power Ratio, is based on the Operational Lethality Indices (OLI) of the units (Note 7).

TABLE 1

COMBAT VARIABLES SELECTED FOR THIS ANALYSIS

1. Duration of the engagement (days)
2. Attacker (1, U.S.; 2, British; 3, German)
3. Winner (0, ambiguous; 1, defender; 2, attacker)
4. Defense posture (1, delay; 2, hasty defense; 3, prepared position, 4, fortified position)
5. Terrain (1, flat; 4, rolling; 7, rugged)
6. Weather (0, clear; 1, light or intermittent rain; 2, rain)
7. Season (1, spring; 2, summer, 3, fall; 4, winter)
8. Attacker initial force (number of men)
9. Attacker casualties (number of men)
10. Defender initial force (number of men)
11. Defender casualties (number of men)
12. Combat power ratio (attacker/defender)

The general characteristics of the engagements are briefly described. Of the 60, there were 19 attacks by British forces, 28 by U.S. forces, and 13 by German forces. The attacker was successful in 34 cases; the defender, in 23; and the outcomes of 3 were ambiguous. With respect to terrain, 19 engagements occurred in flat terrain; 24 in rolling, or intermediate, terrain; and 17 in rugged, or difficult, terrain. Clear weather prevailed in 40 cases; 13 engagements were fought in light or intermittent rain; and 7 in medium or heavy rain. There were 28 spring and summer engagements and 32 fall and winter engagements.

Comparison of World War II Engagements With Historical Battles

Since one purpose of this work is to repeat, in part, Willard's analysis, comparison of these World War II engagements with the historical battles (1618-1905) studied by him will be useful. Table 2 shows a comparison of the distribution of battles by type. Willard's cases were divided into two categories: I. meeting engagements, and II. sieges, attacks on forts, and similar operations. HERO's World War II engagements were divided into four types based on the posture of the defender: 1. delay, 2. hasty defense, 3. prepared position, and 4. fortified position. If postures 1 and 2 are considered very roughly equivalent to Willard's category I, then in both data sets the division into the two gross categories is approximately even.

The distribution of engagements across force ratios, given in Table 3, indicated some differences. Willard's engagements tend to cluster at the lower end of the scale (1-2) and at the

TABLE 2

NUMBERS OF ENGAGEMENTS BY ENGAGEMENT TYPE

Willard (1618-1905)[a]		HERO (1943-1944)	
Type of Engagement	Number	Defense Posture	Number
Category I: Meeting engagements	405	Delay	9
		Hasty Defense	17
Category II: Sieges, attacks on forts, etc.	1088	Prepared Position	14
		Fortified position	20
Totals	1493		60

[a]Table 1, p. 10 (Note 3).

higher end (4 and above), while the majority of the World War II engagements were found in mid-range (1.5 - 4) (Note 8).

The frequency with which the numerically inferior force achieved victory is shown in Table 4. It is seen that in neither data set are force ratios good predictors of success in battle (Note 9).

Results of the Analysis

Willard's Correlation Analysis

There are two forms of the Lanchester equations. One represents the case in which firing units on both sides know the locations of their opponents and can shift their fire to a new target when a "kill" is achieved. This leads to the "square" law where the loss rate is proportional to the opponent's size. The second form represents that situation in which only the general location of the opponent is known. This leads to the "linear" law in which the loss rate is proportional to the product of both force sizes.

As Willard points out, large battles are made up of many smaller fights. Some of these obey one law while others obey the other, so that the overall result should be a combination of the two. Starting with a general formulation of Lanchester's equations, where γ is the exponent of the target unit's size (that is, γ is 0 for the square law and 1 for the linear law), he derives the following linear equation:

$$\log (n_c/m_c) = \log E + \gamma \log (m_0/n_0) \qquad (1)$$

where n_c and m_c are the casualties, E is related to the exchange ratio, and m_0 and n_0 are the initial force sizes. Linear regression produces a value for γ. However, instead of lying between 0 and 1, as expected, the γ's range from -.27 to -.87, with the majority lying around -.5. (Willard obtains several values for γ by dividing his data base in various ways–by force ratio, by casualty ratio, by historical period, and so forth.) A negative γ value is unpleasant. As Willard notes:

> Military theorists should be disconcerted to find $\gamma < 0$, for in this range the results

TABLE 3

NUMBERS OF ENGAGEMENTS BY FORCE RATIO

Willard[a] (1618-1905)

Force Ratio[b]	Number	Cumulative Percent
0.0 - 1.0	--	--
1.0 - 1.5	537	36.0
1.5 - 2.0	302	56.2
2.0 - 2.5	173	67.8
2.5 - 3.0	96	74.2
3 - 4	125	82.6
4 - 5	67	87.0
5 - 6	47	90.2
6 - 7	35	92.5
7	111	100.0
Total	1493	

HERO (1943-1944)

	Number	Cumulative Percent
	3	5.0
	9	20.0
	13	41.0
	15	66.7
	11	85.0
	7	96.7
	0	--
	1	98.3
	1	100.0
	--	
Total	60	

[a] Computed from Table 2, p. 15 (Note 3).

[b] Willard: Force Ratio = (Larger force/ Smaller force)

HERO: Force Ratio = (Attacker/Defender)

TABLE 4

FRACTIONS OF BATTLES WON BY THE NUMERICALLY WEAKER SIDE

Force Ratio[b]	Willard[a] (1618-1905) Category I Fraction	Category I Number	Category II Fraction	Category II Number	HERO (1943-1944) All Engagements Fraction	All Engagements Number
0.0 - 1.0	--	--	--	--	0	3
1.0 - 1.5	.42	473	.39	64	.56	9
1.5 - 2.0	.35	251	.28	51	.23	13
2.0 - 2.5	.25	122	.18	51	.36	15
2.5 - 3.0	.43	58	.13	38	.43	11
3 - 4	.27	56	.25	69	.43	7
4 - 5	.37	30	.19	37	--	0
5 - 6	.11	9	.05	38	1.0	1
6 - 7	.23	13	.05	22	0	1
7	.33	9	.09	102	--	0
	Totals	1,021		472		60

[a]Table 3, p. 15 (Note 3).

[b]Willard: Force Ratio = (Larger force/Smaller force)
HERO: Force Ratio = (Attacker/Defender)

> seem to imply that if the Lanchester formulation is valid, the casualty-producing power of troops increases as they suffer casualties (Note 3).

From his results, Willard concludes that his analysis does not justify the use of Lanchester equations in large-scale situations (Note 10).

Analysis of the World War II Engagements

Willard's computations were repeated for the HERO data set. For these engagements, regression produced a value of -.594 for γ (Note 11), in striking agreement with Willard's results. Following his reasoning would lead to the conclusion that either the Lanchester equations do not represent these engagements, or that the casualty-producing power of forces increases as their size decreases.

However, since the Lanchester equations are so convenient analytically and their use is so widespread, it appeared worthwhile to reconsider this conclusion. In deriving equation (1), Willard used binomial expansions in which he retained only the leading terms. It seemed possible that the poor results might be due, in part, to this approximation. If the first two terms of these expansions are retained, the following equation results:

$$\log (n_c/m_c) = \log E + \log (M_o\text{-}m_c)/(n_o\text{-}n_c) \quad (2)$$

Repeating this regression on the basis of this equation leads to γ = -.413 (Note 12), hardly an improvement over the initial results.

A second attempt was made to salvage

this approach. Starting with raw OLI scores (Note 7), HERO analysts have computed "combat potentials" for both sides in these engagements, taking into account the operational factors of posture, vulnerability, and mobility; environmental factors like weather, season, and terrain; and (when the record warrants) psychological factors like troop training, morale, and the quality of leadership. Replacing the factor (m_o/n_o) in Equation (1) by the combat power ratio produces the result $\gamma = .466$ (Note 13).

While this is an apparent improvement in the value of γ, it is achieved at the expense of somewhat distorting the Lanchester concept. It does preserve the functional form of the equations, but it requires a somewhat strange definition of "killing rates."

Analysis Based on the Differential Lanchester Equations

Analysis of the type carried out by Willard appears to produce very poor results for these World War II engagements. Part of the reason for this is apparent from Figure 1, which shows the scatterplot of the dependent variable, log (n_c/m_c), against the independent variable, log (m_o/n_o). It is clear that no straight line will fit these data very well, and one with a positive slope would not be much worse than the "best" line found by regression. To expect the exponent to account for the wide variation in these data seems unreasonable.

Here, a simpler approach will be taken. Rather than use the data to attempt to discriminate directly between the square and the linear laws, they will be used to estimate linear coefficients under each assumption in turn, starting with the differential formulation rather than the integrated equations used by Willard.

In their simplest differential form, the Lanchester equations may be written:

Square Law: $dA/dt = -k_d D$ and $dD/dt = k_a A$ (3)

Linear Law: $dA/dt = -k'_d AD$ and $dD/dt = k'_a AD$ (4)

where A(D) is the size of the attacker (defender)
dA/dt (dD/dt) is the attacker's (defender's) loss rate,
k_a, k'_a (k_d, k'_d) are the attacker's (defender's) killing rates

For this analysis, the day is taken as the basic time unit, and the loss rate per day is approximated by the casualties per day.

Results of the linear regressions are given in Table 5. No conclusions should be drawn from

TABLE 5

RESULTS FROM THE LINEAR REGRESSIONS: THE 60 WORLD WAR II ENGAGEMENTS

Case	Dep. Var.[a]	Indep. Var.	Corr. Coeff.	Intercept	Slope
Square Law	C_A	D_o	.286	94.64	.01272
	C_D	A_o	.202	54.18	.00707
Linear Law	C_A	D_oA_o[b]	.370	117.44	.00549
	C_D	D_oA_o	.330	75.60	.00627

[a] C_A(C_D) Attacker (defender) casualties per day.
A_o(D_o) Attacker (defender) initial strength.

[b] This product is scaled by 10^{-4} for the regression.

the fact that the correlation coefficients are higher in the linear law case since this is expected for purely technical reasons (Note 14). A better picture of the relationships is again provided by the scatterplots in Figure 2. It is clear from these plots that, as in the case of the logarithmic forms, a single straight line will not fit the entire set of 60 engagements for either of the dependent variables.

To investigate ways in which the data set might profitably be subdivided for analysis, T-tests of the means of the dependent variable were made for several partitionings of the data set. The results, shown in Table 6, suggest that dividing the engagements by defense posture might prove worthwhile.

Results of the linear regressions by defense posture are shown in Table 7. For each posture, the equation that seemed to give a better fit to the data is underlined (Note 15). From this table, the following very tentative conclusions might be drawn:

- In an attack on a fortified position, the attacker suffers casualties by the square law; the defender suffers casualties by the linear law. That is, the defender is aware of the attacker's position, while the attacker knows only the general location of the defender. (This is similar to Deitchman's guerrilla model. Note 16).

- This situation is apparently reversed in the cases of attacks on prepared positions and hasty defenses.

- Delaying situations seem to be treated better by the square law for both attacker and defender.

Table 8 summarizes the killing rates by defense posture. The defender has a much higher killing rate than the attacker (almost 3 to 1) in a fortified position. In a prepared position and hasty defense, the attacker appears to have the advantage. However, in a delaying action, the defender's killing rate is again greater than the attacker's (Note 17).

Figure 3 shows the scatterplots for these cases. Examination of these plots suggests that a tentative answer to the study question posed above might be: "Yes, casualties do appear to be related to the force sizes, but the relationship may not be a simple linear one."

In several of these plots it appears that two or more functional forms may be involved. Consider, for example, the defender's casualties as a function of the attacker's initial strength in the case of a hasty defense. This plot is repeated in Figure 4, where the points appear to fit the curves sketched there. It would appear that there are at least two, possibly three, separate relationships. Also on that plot, the individual engagements have been identified, and it is interesting to note that on the curve marked (1), five of the seven attacks were made by Germans—four of them from the Salerno campaign. It would appear from this that German attacks are associated with higher than average defender casualties for the attacking force size. Since there are so few data points, this cannot be more than a hint or interesting suggestion.

Future Research

This work suggests two conclusions that might have an impact on future lines of research on combat dynamics:

- Tactics appear to be an important determinant of combat results. This conclusion, in itself, does not appear startling, at least not to the military. However, it does not always seem to have been the case that tactical questions have been considered seriously by analysts in their studies of the effects of varying force levels and force mixes.

- Historical data of this type offer rich opportunities for studying the effects of tactics. For example, consideration of the narrative accounts of these battles might permit re-coding the engagements into a larger, more sensitive set of engagement categories. (It would, of course, then be

TABLE 6

T-TESTS FOR SEVERAL PARTITIONINGS OF THE ENGAGEMENTS

		Attacker Casualties Per Day				Defender Casualties Per Day			
	Number Cases	Mean	Standard Error	T-Value	P*	Mean	Standard Error	T-Value	P*
Partitioning by Defense Posture									
Delay	9	111.7	19.0	-1.69	.104	116.4	38.3	-.76	.453
Hasty Defense	17	199.3	36.1	-.57	.571	155.5	31.3	1.58	.124
Prepared Position	14	233.7	49.6	-.10	.921	83.0	33.2	-2.49	.018
Fortified Position	20	240.3	42.9			294.1	66.6		
Partitioning by Terrain									
Flat	19	253.4	40.6	2.00	.054	209.0	40.8	1.28	.208
Rugged	17	148.1	32.3			138.0	36.7		
Partitioning by Season									
Summer, Spring	28	254.2	31.8	2.06	.044	278.4	48.3	3.55	.001
Fall, Winter	32	167.3	27.8			91.9	20.6		

* 2-tail probability

TABLE 7

REGRESSION RESULTS FOR THE ENGAGEMENTS GROUPED BY DEFENSE POSTURE

		Square Law			Linear Law[a]			
Defense Posture	Dependent Variable[b]	Correlation Coefficient	Intercept	Slope[c]	Correlation Coefficient	Intercept	Slope[d]	Number Cases
Fortified	A_c	.430	13.4	.024	.285	148.0	.0051	20
Position	D_c	.023	273.9	.001	.309	138.6	.0087	
Prepared	A_c	.110	180.6	.0064	.436	144.2	.0053	14
Position	D_c	.535	-75.0	.0086	.448	21.5	.0036	
Hasty	A_c	.207	137.5	.0061	.330	125.8	.0044	17
Defense	D_c	.470	-19.7	.0110	.434	71.5	.0050	
Delay	A_c	6.28	-175.0	.0422	.169	71.5	.0330	9
	D_c	.377	-137.3	.0144	.738	-237.0	.0294	

[a] The independent variable was scaled by 10^{-4} for the linear law regressions.

[b] A_c (Attacker) or D_c (Defender) casualties per day.

[c] Interpreted as the killing rate in units of casualties per day per man.

[d] Interpreted as the killing rate in units of casualties per day per man^2 x 10^4.

TABLE 8

KILLING RATES DERIVED FROM
THE REGRESSION EQUATIONS

Engagement Type (Defense Posture)	Killing Rates[a] Attacker	Defender
Fortified Position	.0087	.0240
Prepared Position	.0086	.0053
Hasty Defense	.0110	.0044
Delay	.0144	.0422

[a] It should be noted that the independent variable was scaled by 10^{-4} for the linear law cases. See Table 5.

highly desirable to add more engagements to the data set.)

While predictions of the future are always dangerous, I would nevertheless like to suggest what appears to be a possible trend. While military analysis of the past two decades has focused almost exclusively on the hardware of weapons systems, at least part of our future analysis will be devoted to the more behavioral aspects of combat.

NOTES

1. J.W. Engel, "A Verification of Lanchester's Law," *Operations Research* 2, 163-171 (1954).
2. For example, see R.L. Helmbold, "Some Observations on the Use of Lanchester's Theory for Prediction." *Operations Research* 12, 778-781 (1964); H.K. Weiss, "Lanchester-Type Models of Warfare," Proceedings of the First International Conference on Operational Research, 82-98, ORSA (1957); H.K. Weiss, "Combat Models and Historical Data: The U.S. Civil War," *Operations Research* 14, 750-790 (1966).
3. D. Willard, "Lanchester as a Force in History: An Analysis of Land Battles of the Years 1618-1905," RAC-TD-74, Research Analysis Corporation (1962).
4. The method of computing the killing rates forced a fit at the beginning and end of the battles. See W. Fain, J.B. Fain, L. Feldman, and S. Simon, "Validation of Combat Models Against Historical Data." Professional Paper No. 27, Center for Naval Analyses, Arlington, Virginia (1970).
5. HERO, "A Study of the Relationship of Tactical Air Support Operations to Land Combat, Appendix B, Historical Data Base." Historical Evaluation and Research Organization, report prepared for the Defense Operational Analysis Establishment, U.K.T.S.D., Contract D-4052 (1971).
6. T.N. Dupuy, *The Quantified Judgment Method of*

Analysis of Historical Combat Data, HERO Monograph, (January 1973); HERO, "Statistical Inference in Analysis in Combat," Annex F, *Historical Data Research on Tactical Air Operations*, prepared for Headquarters USAF, Assistant Chief of Staff for Studies and Analysis, Contract No. F-44620-70-C-0058 (1972).

7. The Operational Lethality Index (OLI) is a measure of weapon effectiveness developed by HERO.

8. Since Willard's data did not indicate which side was the attacker, his force ratio is defined to be (larger force/smaller force). The HERO force ratio is (attacker/defender).

9. Since the criteria for success may have been rather different for the two sets of battles, this comparison may not be very meaningful.

10. This work includes more complex analysis in which the possiblity that the two forces may be engaging in different types of combat is considered, leading to the use of two exponents rather than the single one, γ. Stochastic combat processes are also treated.

11. Correlation coefficient = -.262;
Intercept = .00115; slope = -.594.

12. Correlation coefficient = -.184;
Intercept = .0539; slope = -.413.

13. Correlation coefficient = .303;
Intercept = -.638; slope = .466.

14. Correlation coefficients for the linear law are inflated with respect to the square law since the independent variable is a product of force sizes and, thus, has a higher variance than the single force size unit in the square law case.

15. This is a subjective judgment based on the following considerations. Since the correlation coefficient is inflated for the linear law, when it is lower the square law case is chosen. When the linear law correlation coefficient is higher, the case with the intercept closer to 0 is chosen.

16. S.J. Deitchman, "A Lanchester Model of Guerrilla Warfare," *Operations Research* 10, 818-812 (1962).

17. As pointed out by Mr. Alan Washburn, who prepared a critique on this paper, when comparing numerical values of the square law and linear law killing rates, the differences in units must be considered. (See footnotes to Table 7).

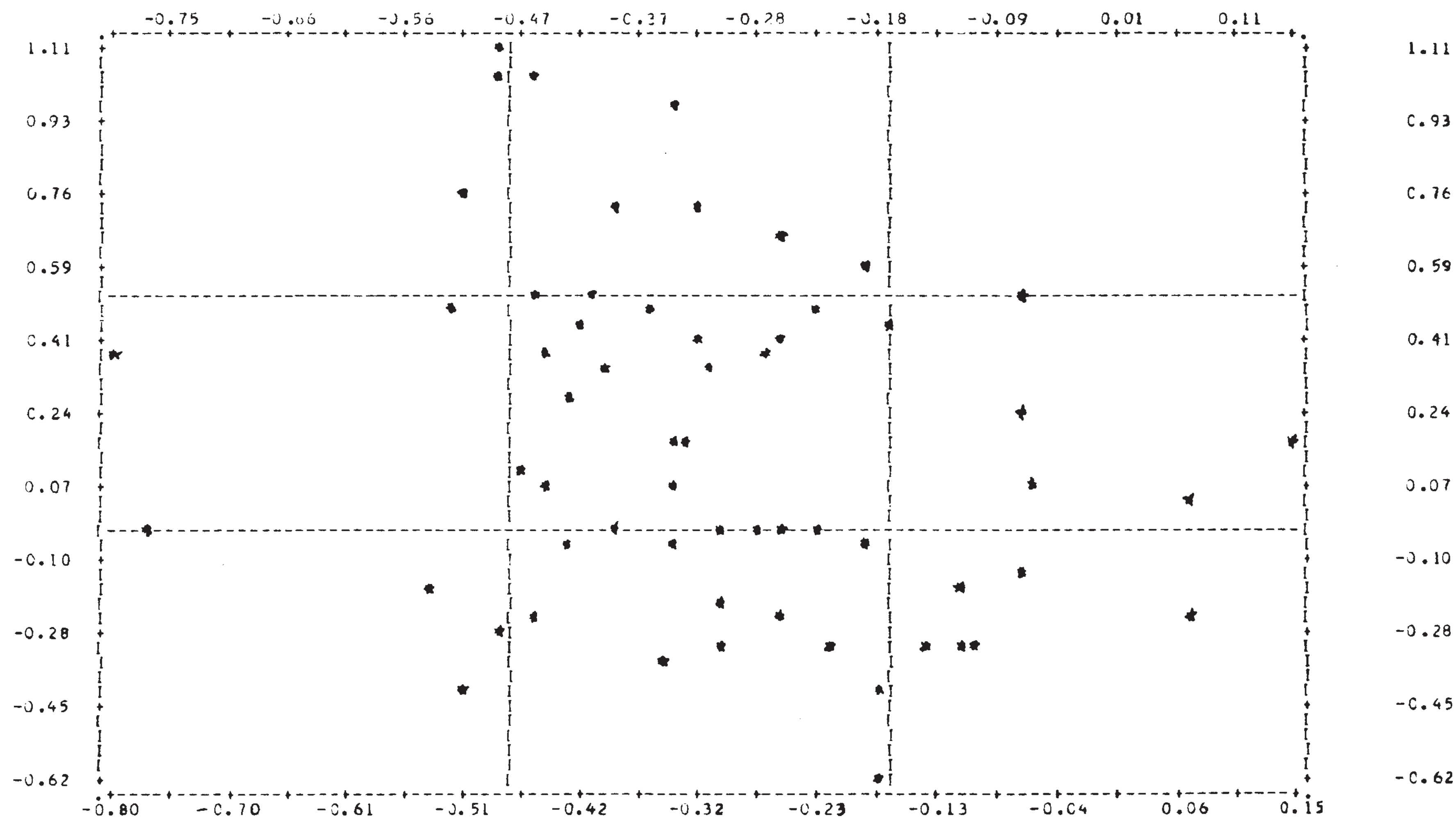
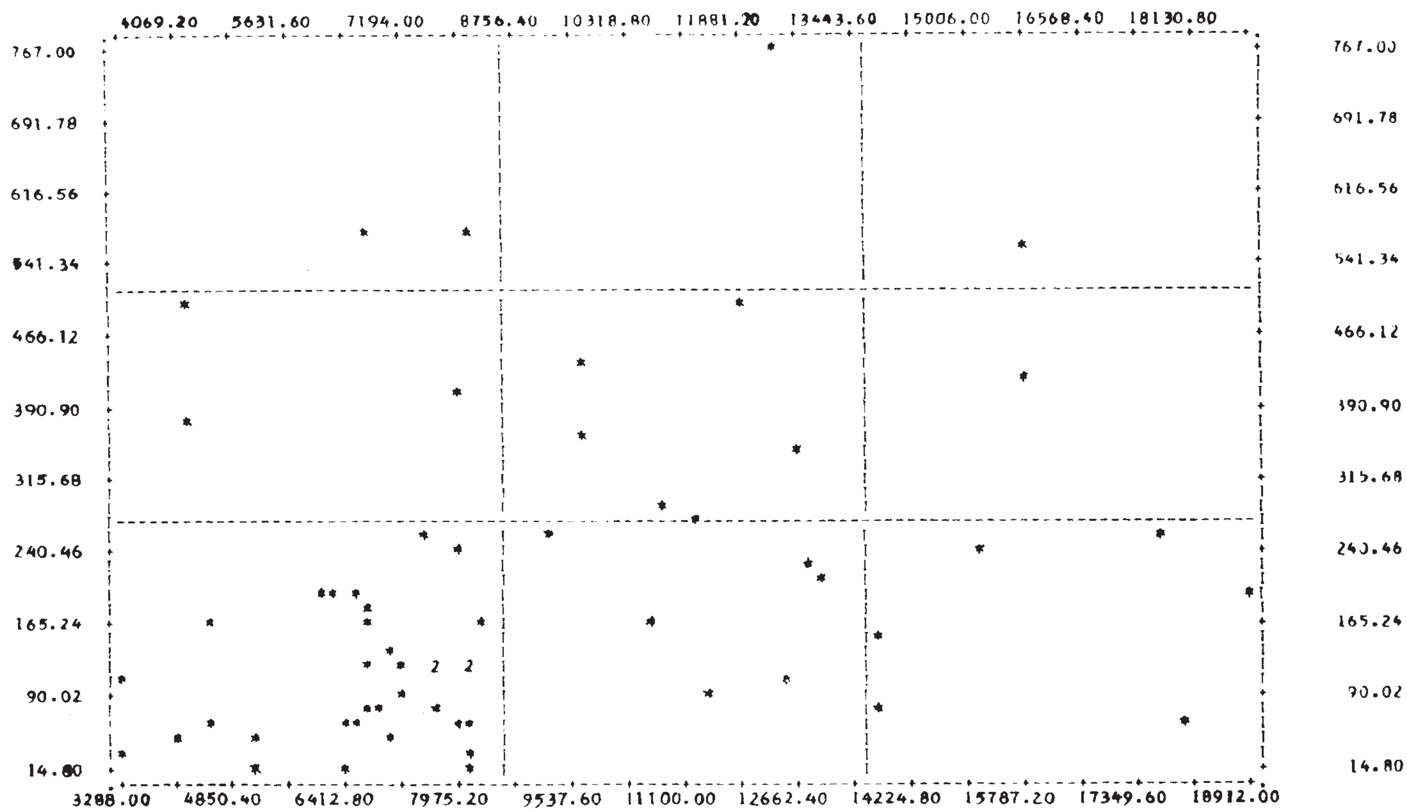


Figure 1. Scattergram for 60 World War II Engagements
Y: Dependent Variable: LOG_{10} (Attacker Casualties/Defender Casualties)
X: Independent Variable: LOG_{10} (Defender Initial Strength/Attacker Initial Strength)

Dependent Variable: Attacker Casualties

Square Law



Linear Law

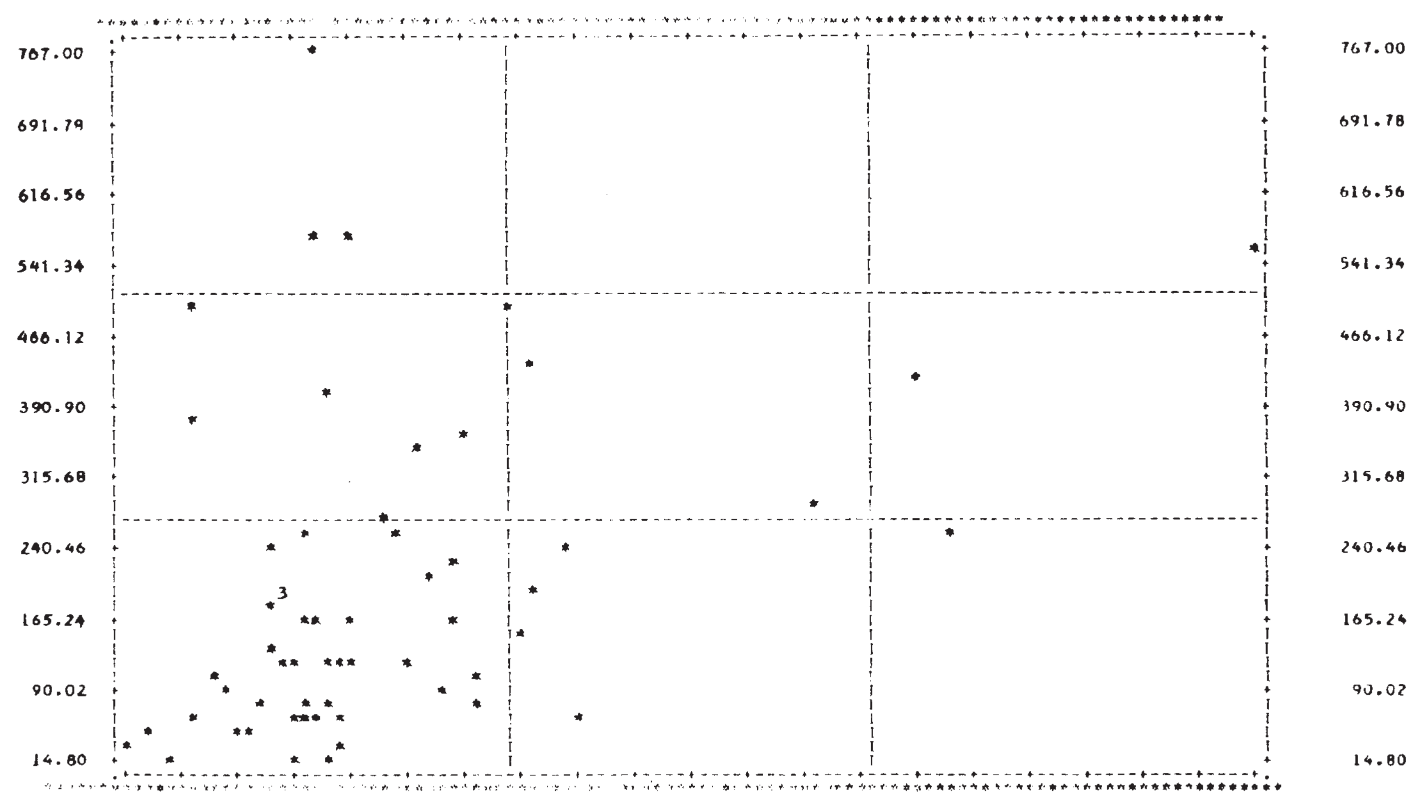
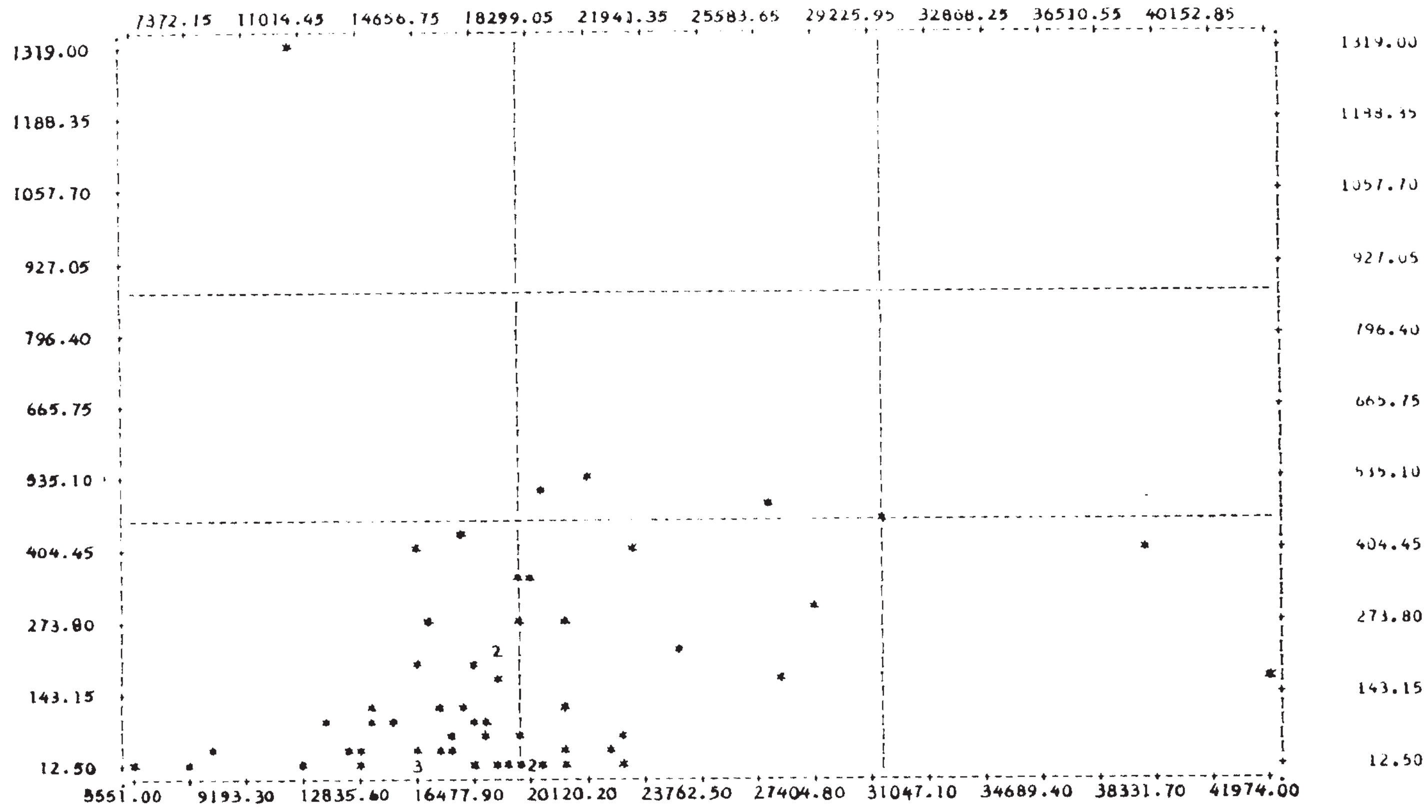


Figure 2. Scattergrams for 60 World War II Engagements

Dependent Variable: Defender Casualties

Square Law



Linear Law

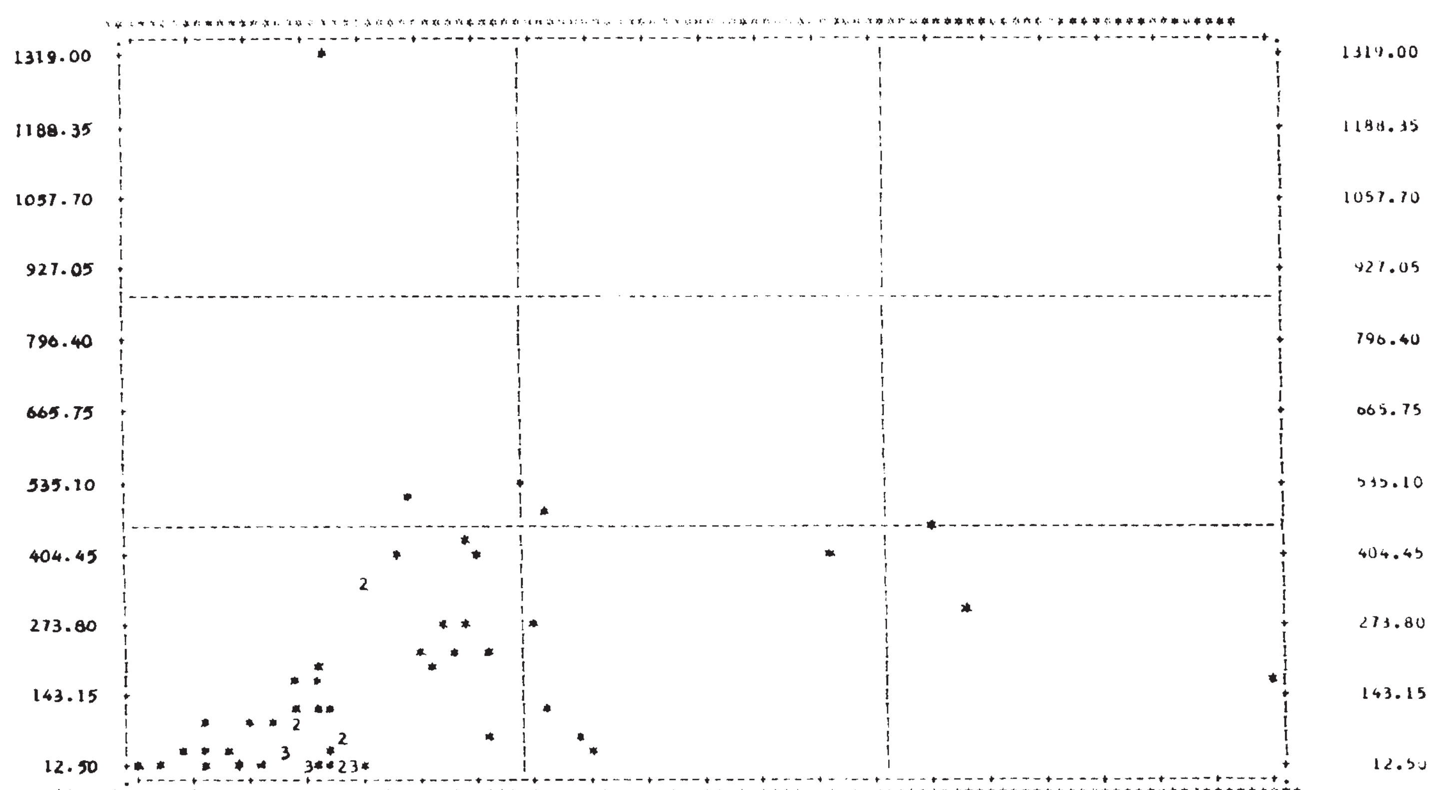
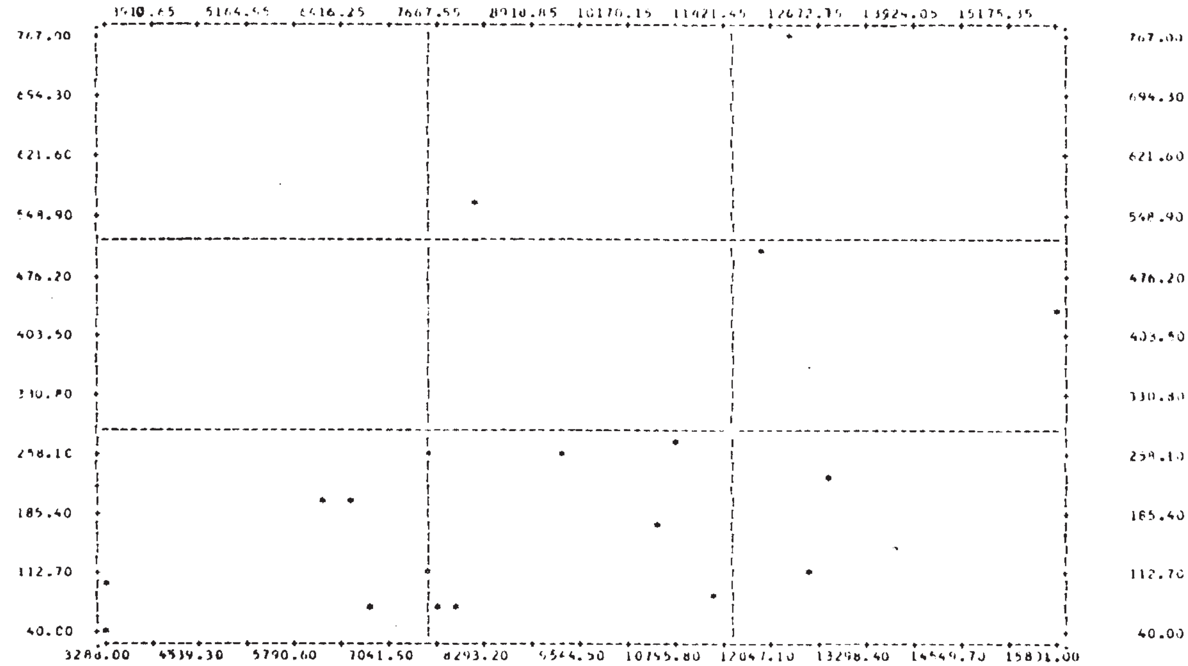


Figure 2. (Continued)

FORTIFIED POSITION

Attacker: Square Law



Defender: Linear Law

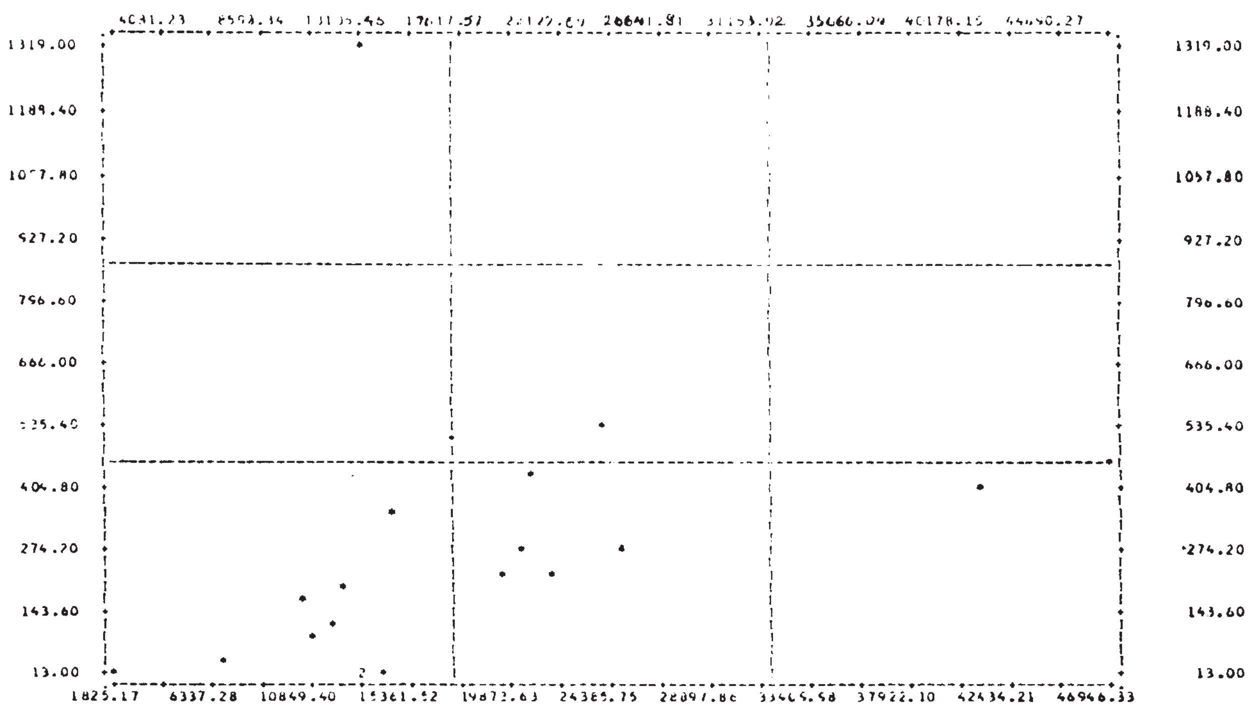
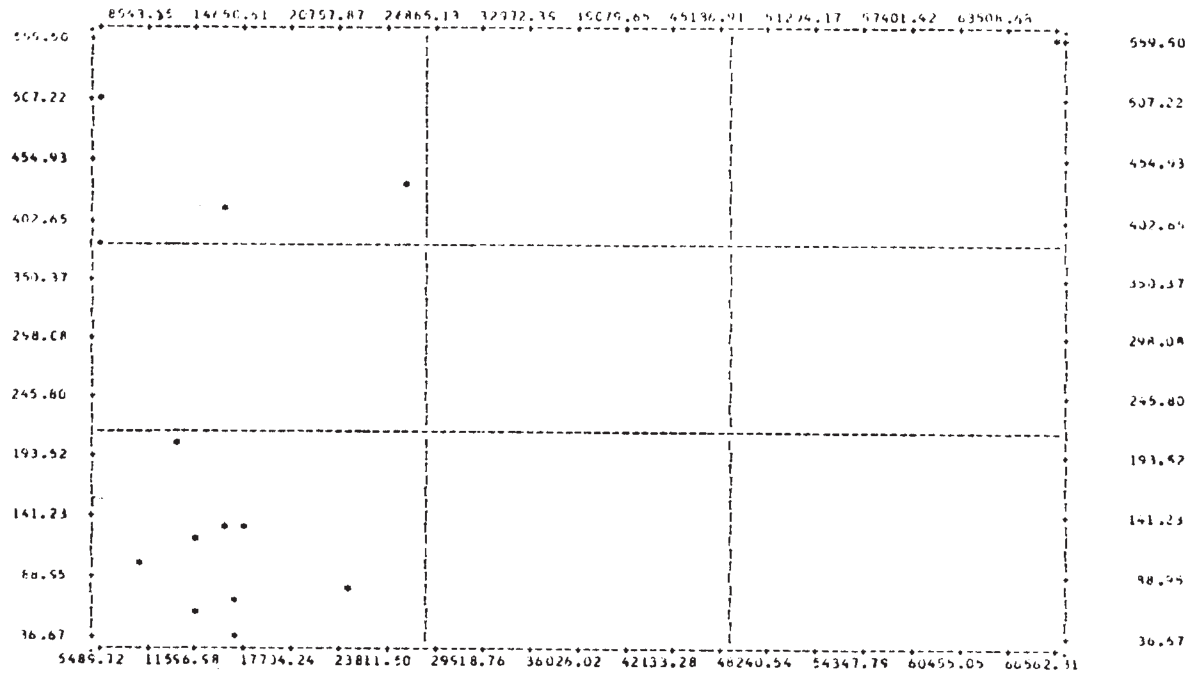


Figure 3. Scattergrams for Attacker and Defender by Defense Position

PREPARED POSITION

Attacker: Linear Law



Defender: Square Law

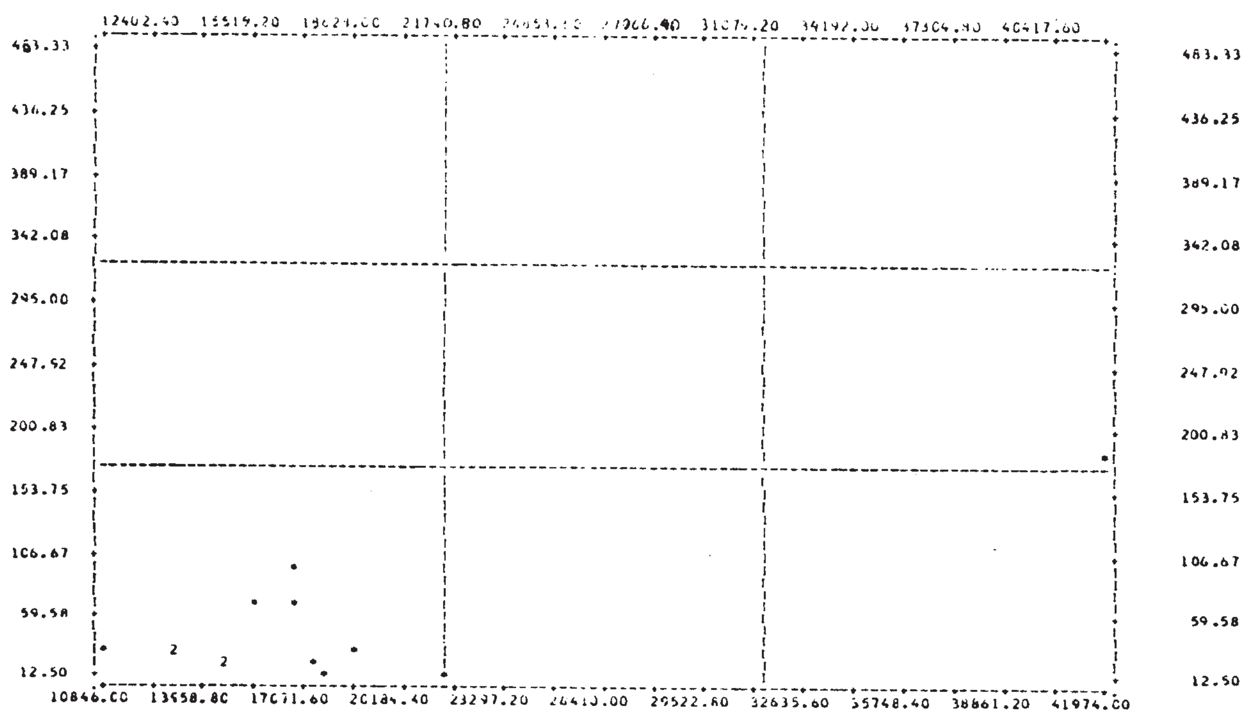
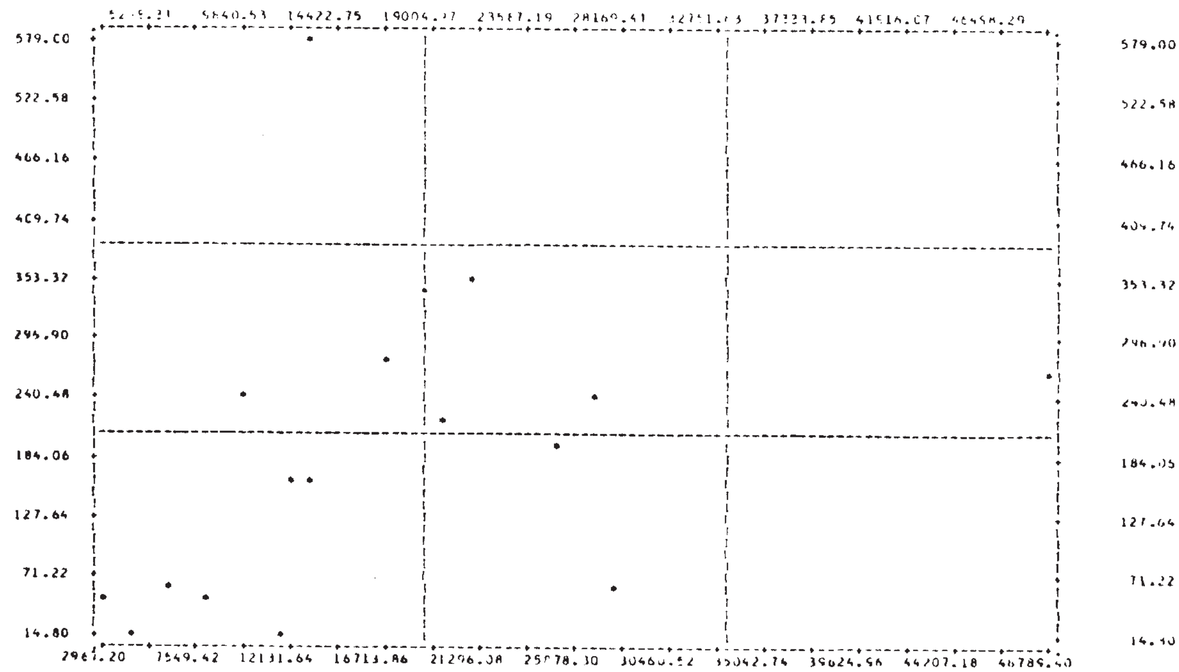


Figure 3. (Continued)

HASTY DEFENSE

Attacker: Linear Law



Defender: Square Law

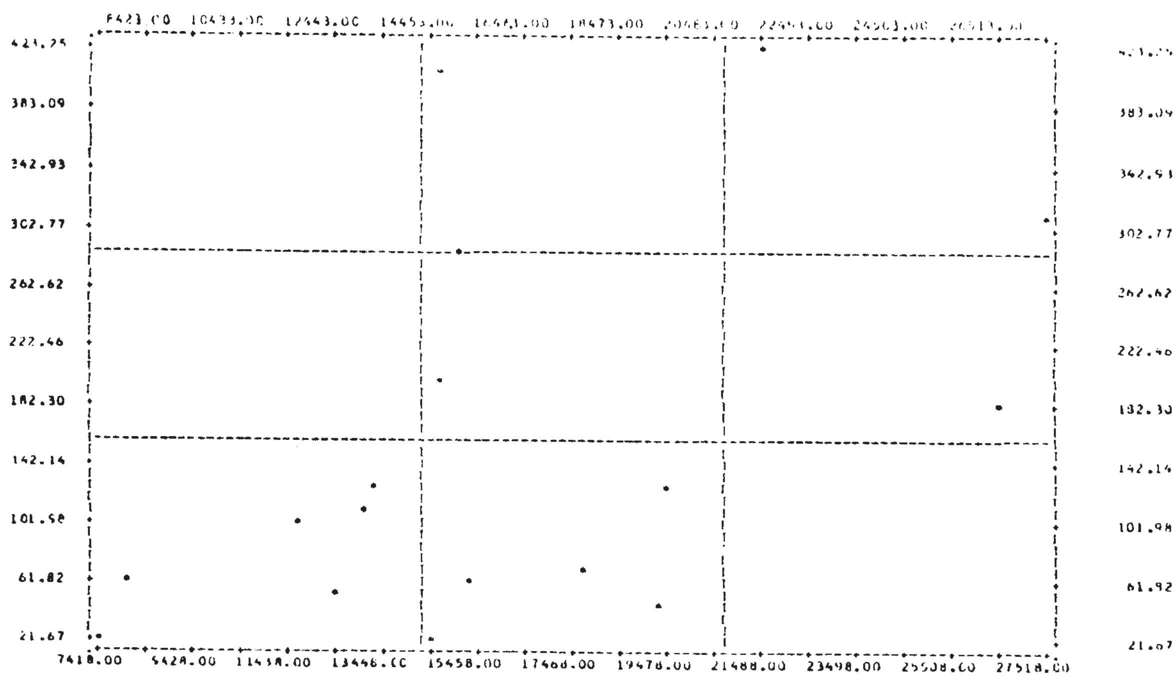
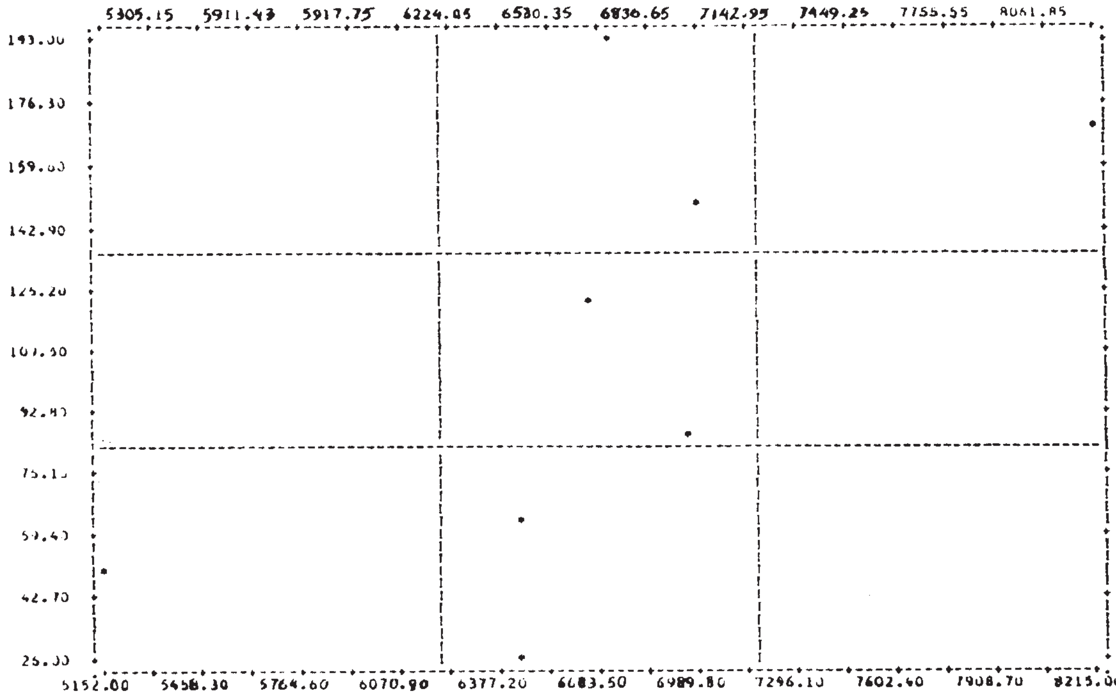


Figure 3. (Continued)

DELAY

Attacker: Square Law



Defender: Square Law

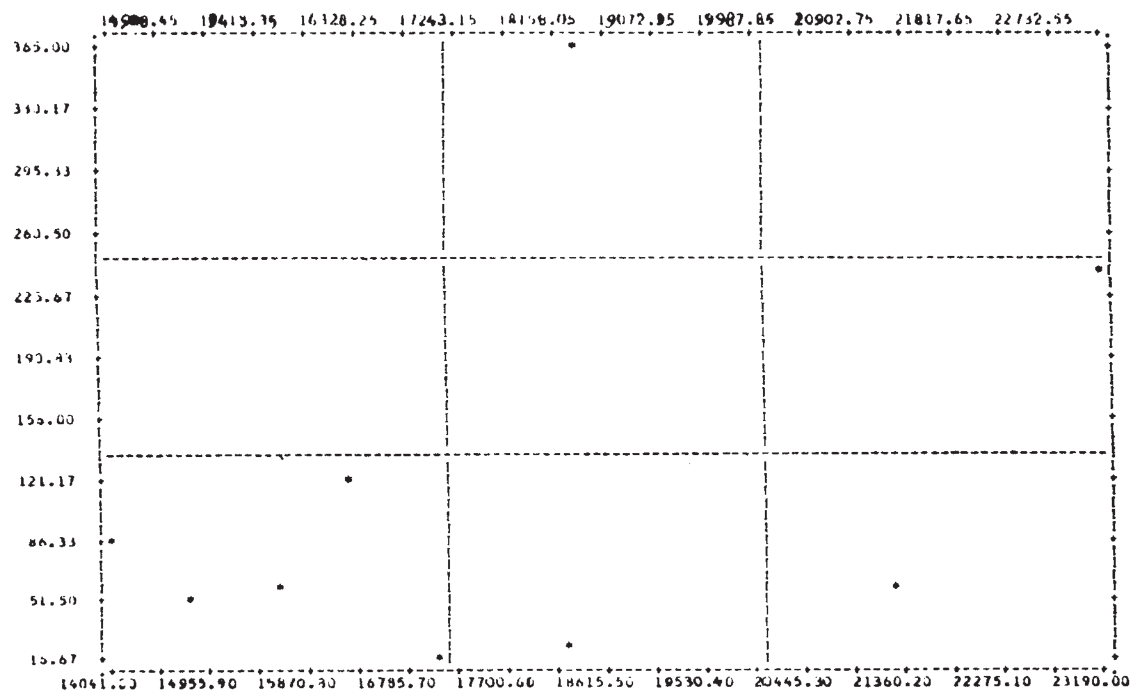


Figure 3. (Continued)

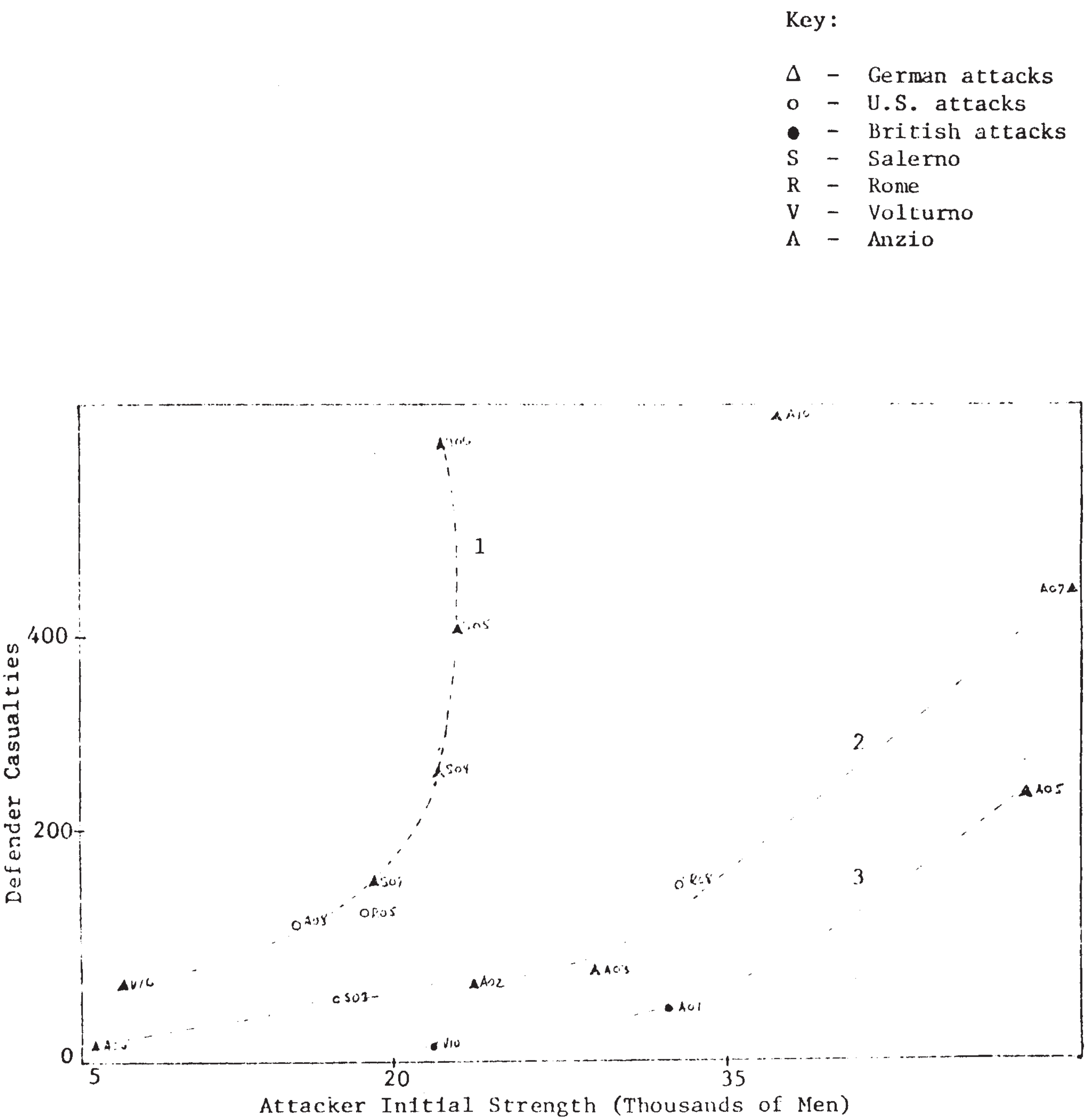


Figure 4. Defender Casualties Versus Attacker Initial Strength for Hasty Defense