Uploaded to VFC Website
~ October 2012 ~

This Document has been provided to you courtesy of Veterans-For-Change!

Feel free to pass to any veteran who might be able to use this information!

For thousands more files like this and hundreds of links to useful information, and hundreds of "Frequently Asked Questions, please go to:

Veterans-For-Change

Veterans-For-Change is a 501(c)(3) Non-Profit Corporation
Tax ID #27-3820181

If Veteran's don't help Veteran's, who will?

We appreciate all donations to continue to provide information and services to Veterans and their families.

https://www.paypal.com/cgi-bin/webscr?cmd=_s-xclick&hosted_button_id=WGT2M5UTB9A78

Note: VFC is not liable for source information in this document, it is merely provided as a courtesy to our members.

Item ID Number 01735

Author

Corporate Author Agent Orange Projects, Chronic Disease Division, Cente

Report/Article Title Typescript: Agent Orange Projects Interim Report (Draft): Agent Orange Study, Exposure Assessment: Procedures and Statistical Issues, February 1985

Journal/Book Title

Year 0000

Month/Day

Color ☐

Number of Images 27

Descripton Notes

DRAFT

Agent Orange Projects Interim Report

AGENT ORANGE STUDY

EXPOSURE ASSESSMENT: PROCEDURES AND STATISTICAL ISSUES

February 1985

Agent Orange Projects

Chronic Diseases Division

Center for Environmental Health

Centers for Disease Control

TABLE OF CONTENTS

I. SUMMARY

Because of the critical importance of exposure assessment in the Agent Orange Study, this Interim Report presents an update of work on several related issues since the approved protocol for CDC's Epidemiologic Studies of the Health of Vietnam veterans was published in November, 1983.

In Section III several needed modifications of procedures for identification and selection of men for the study cohorts are described. Although not strictly part of exposure assessment, cohort selection procedures can influence both the likelihood of Agent Orange exposure in study participants and the methods to estimate it. In turn, cohort selection procedures are influenced by the availability of the unit daily location data which are used in the assessment of exposure. Four major changes from the methods of cohort selection described in the protocol are necessary on the basis of several pretests of methods for the study: 1) Battalions, rather than companies, will serve as the basis for cohort selection and unit location. 2) Ranking of unit exposure likelihoods will not determine selection of men into the study. 3) Individual military personnel records, as well unit morning reports, will be used to identify veterans period of service in Vietnam. 4) The study will be restricted to men serving in infantry or artillery units.

The two most important impacts of these changes on the study design are: First, because battalions are larger, more dispersed units than companies, the location of men in relation to herbicide applications will be known with less precision than envisioned in the protocol, and Second, because unit exposure likelihoods do not appear to correlate highly with individual men's exposure likelihoods, we will not be able to exclude men from units with intermediate exposure likelihoods from the study; therefore the spread or range of exposure likelihoods between the groups being compared will likely be reduced.

Section IV describes the sources and some characteristics of the data on battalion daily locations and herbicide applications that will be combined to assess exposure likelihood. Abstraction of battalion daily location by ESG involves a hierarchical search of the available military records for location information. If the preferred battalion records are not available, then higher level records are used (brigade, division, and field force, in that order). Within each level the abstractor first searches the daily journal and if it is not adequate , the situation reports, ORLLs (Operations Report Lessons Learned), COAARs (Combat Operation After Action Report), command reports, and intelligence summaries, in that order. Data on herbicide applications are available from two computerized sources: the Herbs tape which Ranch Hand applications by fixed wing and, after 1968 rotary winged aircraft and the Services Herbs tape which provides information or other applications, such as helicopter or jeep based sprayings, leaks and other ground sprayings.

Section V discusses two mathematical models for assessing the likelihood of Agent Orange exposure, the exponential-decay model and the "hits" model. This section also mentions possible values for the time and distance parameters in the models, as well as the actual procedures involved in ranking veterans' likelihood of exposure.

The misclassification of Agent Orange exposure inherent in this study means that the true magnitude of exposure-disease associations will be underestimated. Section VI discusses this topic, as well as some of our preliminary efforts to quantify the influence of various sources of misclassification.

We welcome advice and comments on these issues.

II. INTRODUCTION

The Agent Orange Study is one component of the Centers for Disease Control's congressionally mandated Epidemiologic Study of the Health of Vietnam Veterans. The purpose of this component is to determine whether U.S. Army combat troops who served near where the herbicide Agent Orange was sprayed in Vietnam are at an increased risk of death, disease, or disability, compared to troops who served further from the spray missions.

This report presents information on several issues related to assessment of the likelihood of exposure of men in the Agent Orange Study. The first section reviews the protocol's description of operational methods for the selection and identification of men for the study cohorts, particularly military records abstraction procedures. Data from pretests of these procedures indicate that some should be modified. The remainder of the report discusses the data on daily troop locations and herbicide spraying missions that will be used to rank individual veteran's likelihood of exposure to Agent Orange, results of initial research concerning statistical models for ranking the likelihood of exposure to Agent Orange and some of the effects and causes of misclassification of exposure in this study.

II. COHORT SELECTION PROCEDURES

A. Background

The Agent Orange Study involves three cohorts, each containing approximately 8,500 men. All will be followed for mortality status We estimate that at least 6,000 men in each cohort will agree to be interviewed; 2,000 of those interviewed will receive a complete medical examination. Two of the cohorts will consist of Army veterans who served in those areas of Vietnam receiving the heaviest administrations of herbicides (III Corps). The third cohort will be comprised of Army men from service support units stationed in areas where there were almost no herbicide sprayings. All of the participants must also meet the following criteria: single enlistment, rank of E-5 or under at discharge, and a single tour of duty in Vietnam entirely during 1967-1968.

B. Modifications

Section 4.1.1. of the protocol describes the related processes of selecting men for the three cohorts, abstracting military records containing information on the daily location of the men's units, and ranking the men's likelihood of exposure to Agent Orange based on the daily proximity of their units to herbicide spraying missions. The protocol notes that results of pretests of the procedures might necessitate their modification. Based on our research, and that of the U.S. Army and Joint Services Environmental Support Group (ESG), several of the procedures described in the protocol will, in fact, need to be modified.

The most important changes will occur in the process of selection of men into the cohorts: 1)Battalions, rather than companies, will serve as the basis for cohort selection and unit location. 2) Ranking of unit exposure likelihoods will not determine selection of men into the study. 3) Individual military personnel records, as well unit morning reports, will be used to identify veterans period of service in Vietnam. 4) The study will be restricted to men serving in infantry or artillery units.

1) Battalions - rather than companies - will serve as the basis for cohort selection and unit location

In the protocol, an early step in the selection of men for the Agent Orange Study was to ascertain the geographic location of company- or battery-sized units from selected battalions in III Corps for each day of the 731-day period from 1-1-67 until 12-31-68. Current data indicate that daily location information is not consistently available for units smaller than battalions. In other words, we believe that the location of individual battalions can be placed with reasonable accuracy for most of the 731 days but the locations of the component companies of a battalion cannot be as reliably placed due to lack of data.

Table 1 summarizes data on the availability of daily unit coordinates for 21 battalions reviewed by ESG through December 31, 1984. The table presents the number of days for which at least one point is available for each battalion and each company within the battalions as well as the total number of abstracted points for each unit over the entire two years. The battalion numbers are from the AOP master list (supplied by ESG) of 65 battalions serving in III Corps for at least 18 months in 1967 and 1968. The table demonstrates that the data are incomplete for both battalion- and company-level locations.

Two main opportunities exist for filling these "gaps", or days for which no geographic coordinates are available: 1) Supplement review of battalion records with review of division and brigade records. These higher level records are more likely to contain coordinates on battalions than on smaller units such as companies or batteries. 2) Utilize data for days on which a unit's geographic location is described by a name or name code instead of actual grid coordinates. By a procedure that will be described in Section IV.A.2 of this report, many of these name codes can be linked to the coordinates to which they refer.

A recent pretest performed by ESG and evaluated by AOP looked at the success of division and brigade record review and name code - coordinate linkage in filling gaps on 9 of the 21 battalions listed in Table 1. The results indicate that sufficient information is available on battalions, but not companies, to fulfill the protocol's requirement of no more than 30 contiguous days of missing records nor more than 60 days missing during the entire two year period. For these reasons battalions rather than companies or batteries will be the units whose locations form the basis for ranking individual men's likelihood of exposure to Agent Orange. Because battalions are larger and generally more dispersed than companies, the location of men in relation to spray missions will be known with less precision than envisioned in the protocol. This issue is discussed more fully in Sections IV.A. and VI.B.3.

Table 1
Number of individual coordinate points noted and the number of days for which at least one location is known for battalions and companies within battalions. III Corps, January 1, 1967 - December 31, 1968.

Bat.	COMPANY										TOTAL	
	A		B		C		D		E		BATTALION	
#	PTS	DAYS	PTS	DAYS	PTS	DAYS	PTS	DAYS	PTS	DAYS	PTS	DAYS
1.	663	190	717	214	656	196	499	140			4235	479
2.	1498	312	1575	401	1537	433	94	46	32	21	9663	695
3.	2021	417	2098	431	2209	424	1477	232	772	299	9243	650
4.	1801	338	1914	352	1661	319	14	13	2	2	8929	477
5.	891	163	386	69	550	81					5054	479
6.	3718	702	1175	641	1456	650	812	536	2	1	7552	706
7.	119	64	116	68	36	18	106	50	8	4	955	450
8.	67	35	36	28	76	47	31	24	2		280	125
9.	345	263	347	287	352	284	73	69	1	1	1940	424
10.	1739	404	1474	369	1083	224	1366	379	4	2	6880	631
11.	2597	583	2788	551	2708	565	1698	306	1382	485	11812	680
12.	1654	467	1809	478	1657	465	1198	318	1	1	7136	617
13.	1561	416	1645	426	1664	443	1376	257			7441	468
14.	2538	492	2337	452	2563	470	2001	265	1288	365	12218	693
15.	3	1	47	10	19	1	35	2			1134	275
16.	382	246	397	251	424	295					1904	547
18.	34	29	67	60	83	80					883	190
20.	2072	300	2609	362	1973	310	113	49			8748	643
21.	2260	466	2178	466	1868	440	1502	328	59	7	9000	727
24.	1104	369	694	284	875	337	535	243			4259	674
26.	128	45	88	38	104	39	94	42			1293	590

2) Ranking of unit exposure likelihoods will not determine selection of men into the study.

The protocol specified that the first 125 companies (25 battalions) should be ranked according to time and distance proximity to herbicide applications before any men are selected for the study. The process of selecting study participants would then consist of identifying those men who served at least nine months in companies falling in the bottom third and top third of these rankings. No information would have been collected on men who served in companies which comprised the middle third of the rankings. By excluding men from the middle third, we hoped to maximize differences in the likelihood of exposure between the two cohorts from III Corps.

However, this procedure was predicated on the belief that the rankings of companies according to time and distance from herbicide applications would correlate highly with the rankings for individuals. By selecting only men who came from companies in the top and bottom third of the exposure rankings, we expected to accomplish the same goal as if we had collected information on all of the men individually, ranked them according to time and distance from herbicide applications, and then selected only those men in the top and bottom third of these individual exposure rankings. Recent information (discussed in Section III.B.3.) indicates, however, the assumption that the ranking of companies would correlate with the ranking of individuals is incorrect. Many men transferred between battalions within III Corps and therefore their likelihood of exposure would not correlate highly with the likelihood of exposure over the two year interval of any single battalion. In addition, the unexpectedly high transfer rate among companies in III Corps means that not enough men would fulfill the recommended criterion of having spent nine months in a single company to yield the desired cohort size. Therfore, eligibility for inclusion in the study will be based on continuous time stationed in III Corps, rather than in a single unit. We have not determined what the exact time requirement will be, but may use a minimum of nine months service anywhere in III Corps because of its similarity to the original criterion of nine or more months service in a single unit.

The major drawback of the high transfer rate among units in III Corps is that it will prevent the exclusion of men in units in the middle third of battalion exposure likelihood rankings, thereby probably reducing the spread of exposure intensity among the men in the study. Some of the ramifications of this change will be considered in future computer simulations that are mentioned in Section VI. In addition, the high transfer rate necessitates obtaining location information on all the battalions that were located in III Corps for at least 18 months during 1967 and 1968, rather than a sample of these battalions.

3) Individual military personnel records, as well unit morning reports, will be used to identify veterans' period of service in Vietnam

Selection of individuals from companies was to be based primarily on information from company morning reports. These daily reports provide information on all company personnel actions such as assignments, transfers, and promotions. The veteran's name and military service number (MSN) were usually included in the report, along with information concerning the type of personnel action taken. Theoretically, men could be selected who had served at least nine months in these companies by reviewing information on unit arrivals and departures contained in the company-level morning reports. A pretest conducted by ESG indicates, however, that we will not be able to determine the amount of time that individuals spent with particular units by relying solely on the morning reports. The ESG pretest is summarized in Table 2.

Table 2
Duration of Veteran's Assignments to Combat Companies
Morning Report Abstraction Pretest Conducted by ESG
January 1, 1967 - December 31, 1968

Number of Soldiers Identified

	Duration of assignment to company (mos.)					Percent
	12+	9-12	Uncertain	9	Total	uncertain
Company type:						
artillery	14	48	57	200	319	17.9
infantry	4	55	171	427	657	26.0
infantry	0	29	129	423	581	22.2
cavalry	8	78	233	159	478	48.7
artillery	5	42	161	69	277	58.1
artillery	2	48	92	50	192	47.9
artillery	11	59	77	27	174	44.3
infantry	0	64	317	239	620	51.1
engineer	5	70	117	93	285	41.1
artillery	4	59	122	70	255	47.8
Total:						
infantry	4	148	617	1089	1858	33.2
artillery	36	256	509	416	1217	41.8
Total	53	552	1476	1757	3838	38.5
Percent	1.4%	14.4%	38.5%	45.8%	100%	

Note: This table restricted to men of rank E-5 or lower

The pretest used Army morning report records to categorize the men into four groups: those serving 12 or more months in their companies, those serving 9-12 months, those serving less than 9 months, and those for whom the amount of time served in their companies is unknown. The tracking was judged to have failed if a veteran was reported as present in a unit during the specified time period, but the length of time spent with the unit could not be ascertained using available morning reports. The rate of unsuccessful tracking was high (38.5%), and only 15.8% of the men met the eligibility criterion of documented presence in the unit for at least nine months. This low a yield from all the battalions would prevent the identification of enough eligible men in all of III Corps to fill the two cohorts selected from that area. Equally importantly, the low percentage of eligible men based on morning report abstraction raises questions of the "representativeness" of those men who did stay in a single unit for more than 9 months, and have adequate morning report documentation of their arrival and departure from the unit to permit determination of their exact length of time in unit.

The low yield from morning report abstraction appears to be due to three factors: 1) the transfer of men among units alluded to earlier, and 2) the restriction of the pretest to records for 1967 and 1968 - some men may have arrived as early as October 1966 or left as late as April 1969 and still have served 9 months of their year tour during 1967 and 1968, 3) missing morning reports that prevent the determination of exactly when a soldier arrived in or left a unit.

As part of the development of quality control procedures, a team from AOP visited the National Personnel Records Center (NPRC) in St. Louis to abstract information from the military personnel files of 216 men identified in an earlier pretest as serving in an infantry battalion in III Corps during 1968. Because of concerns about transfers among companies and the inability to track men using only military morning reports, AOP studied these men's records in detail.

This pretest found that we can determine in which battalion men served at different times from their personnel files. These files contain information on all transfers between units, as well as more complete information on arrival and departure dates than can be obtained from extant morning report records. Complete information on daily battalion assignments was obtained for 109 of the 116 men (96.8%), a much higher yield than in ESG's pretest based solely on morning reports. One limitation should be noted: we are not able to learn of short-term absences from the unit such as week-end passes and R and R from personnel file records. However, no reliable method is available to establish this information.

In addition, we found that 116 men (53.7%) met all of the study criteria except for the qualification that they have spent at least nine months in one company. However, we also learned that 94 out of 116 men, while not actually serving in a single infantry unit for 9 months, did spend at least 9 months in III Corps stationed in various companies of the 65 battalions identified as being located in III Corps for at least 18 of the 24 months from 1-1-67 to 12-31-68.

Since we cannot acquire accurate information on the daily presence of men in units using morning reports, we can simplify the abstraction of these records. Morning reports, and other unit records such as periodic rosters, will be used only to identify men who served in battalions in III Corps during 1967 and 1968. ESG believes that, although these unit records cannot determine the exact duration of men's service in a battalion, they can accurately identify those men who served in the battalion at some time in 1967 or 1968. All detailed information on the period of service in III Corps and in specific units then will be obtained on the men from their individual personnel files.

4) The study will be restricted to men serving in infantry or artillery units.

According to information from ESG, 67 Army combat battalions were stationed in III Corps for at least 18 months during 1967 and 1968. Of the 65 battalions, 27 were infantry, 25 were artillery, 5 were cavalry, 2 were armor, and 6 were engineering units.

AOP has decided to select only individuals from infantry and artillery units from the Agent Orange Study for two reasons. First, companies from the cavalry, armor, and engineering battalions were widely dispersed. Therefore, the "average" battalion location on any given day would be a poor indicator of the location of any of the constituent companies or batteries. Second, the health of Vietnam veterans who served in calvalry, armor, and engineering units will be studied in the Vietnam Experience Study. For the Agent Orange Study, our goal is not to select a "representative" sample of Army men in III Corps, but rather to select men from III Corps who are as comparable as possible, except for differing likelihoods of Agent Orange exposure. We believe that men serving in artillery and infantry battalions will resemble each other more closely in terms of baseline characteristics and general military experience than they would resemble men from cavalry, armor, and engineering units.

C. Flow Chart of Modified Methods

The modified procedures for military record abstraction and cohort selection are summarized in the flow chart in Figure 1.

IV. EXPOSURE ASSESSMENT DATA

Procedures involved in exposure assessment are also outlined in Figure 1. Key to the process of determining the proximity of men to Agent Orange sprayings is the linkage of data on battalion daily locations with data on the date and location of herbicide applications. This section discusses the availability of these data and mentions operational and methodological issues related to the abstraction and analysis of the data.

A. Battalion Daily Locations

1. Background

Abstraction of battalion daily location by ESG involves a hierarchical search of the available military records for location information. If the preferred battalion records are not available, then higher level records are used (brigade, division, and field force, in that order). Within each level the abstractor first searches the daily journal and if it is not available , the situation reports, ORLLs (Operations Report Lessons Learned), COAARs (Combat Operation After Action Report), command reports, and intelligence summaries, in that order.

The abstractor records any place names and map coordinates describing the location of a battalion unit, the dates these coordinates or place names were entered into the military record, the size of the unit to which the coordinates or place names are attached, and the type of document from which the information is abstracted. The map coordinates are in the Universal Transverse Mercator (UTM) system which uses a two letter and six digit designation of location. The letters refer to a 100 km by 100 km grid on the map of Vietnam, and the first three digits divide the east-west direction into 0.1 km grids, while the second three digits accomplish the same thing for the north-south direction. Unfortunately, adequate information does not exist on all the battalions for each day of the period being studied (see Table 1 above).

2. Place Names

ESG has documented that the only information available concerning a battalion's location on certain days is a place name (e.g. Binh Phouc). A place name may be recorded in the military record with or without associated map coordinates. This information is included on the abstraction form using a uniform set of abbreviations for the place names. Place names, however, are not always linked to a unique set of coordinates. For example, some names have occurred with several sets of coordinates, often located some distance from one another. Therefore we can not simply apply a mechanical algorithm to assign coordinates to name codes occurring without them. Table 3 summarizes the information available from the location data on the 5th Battalion, 2nd Artillery Regiment, II Field Force on coordinates associated with Binh Phouc. This place name is associated with 4 distinct locations that are separated by as much as 100 km.

Table 3
Coordinates Associated with the
Place Name Binh Phouc

Coordinate	Number of Occurrences
XS600500	1
XS138448	1
XT600550	50
XT882012	16

Since place names often are associated with different sets of coordinates, the reasonableness of a particular set relative to the battalion's other reported positions must be the determining factor for deciding location and not the place name. Coordinates will be assigned to those place names that occur without them by a panel of AOP staff who will determine assignment on the basis of context of other known locations for the battalion. All decisions of the panel will be based on consensus, and all imputed coordinates will be flagged in the data set so that the impact of this procedure can be analyzed. When we have finished this process, we hope to establish the location of a battalion for each day since classifying exposure depends on determining the distance from this these daily locations to herbicide applications. Since this procedure has not been thoroughly tested, it may have to be modified, but we hope that this process of imputing locations will provide some information for days on which we presently have none.

3. Location Summarization

For most days a battalion will have several sets of map coordinates associated with it. Table 4 summarizes the data available for the 1st Battalion, 2nd Infantry Regiment, 1st Brigade, 1st Infantry Division for the five day period from 2/6/67 until 2/10/67.

Table 4
Known Locations on Specified Dates for the 1st battalion according to the AOP Master List
8

	02/06/67		02/07/67		02/08/67		02/09/67		02/10/67	
	X	Y	X	Y	X	Y	X	Y	X	Y
1	359.9	364.9	359.9	364.9	358.0	367.3	358.0	366.6	362.9	366.5
2	361.4	360.0	364.0	367.0	358.0	368.0	362.6	364.0	364.0	367.0
3	364.0	367.0	364.0	367.0	358.8	368.7	362.9	366.5	364.4	367.1
4	364.1	363.0	396.0	348.0	358.8	368.2	364.0	367.0	364.4	367.1
5	396.0	348.0	396.5	348.4	362.4	366.8	364.4	367.1	396.0	348.0
6	396.7	348.4	397.4	349.3	362.9	366.5	396.0	348.0	396.8	348.4
7	397.5	349.0	397.4	349.9	364.0	367.0	396.6	348.3	397.3	349.9
8					364.4	367.1	397.4	349.3	397.5	349.0
9					364.4	367.1	397.4	349.9		
10					365.9	369.6				
11					396.0	348.0				
13					396.8	348.3				
14					397.3	349.9				
15					397.5	349.0				
Average	377.1	357.2	382.2	356.4	371.8	362.3	377.7	358.5	380.4	357.9
St.dev.	17.1	7.8	17.0	8.7	16.1	8.6	17.2	8.7	16.5	9.1
Scatter	133.2		146.9		137.4		149.6		149.6	
St.Dist.	18.8		19.1		18.2		19.3		18.8	

6 In this table the UTM coordinates have been transformed into a cartesian coordinate system to simplify in calculating distances and summary statistics. The numbers X and Y determine the distance in kilometers of the point from an arbitrarily chosen origin. This distance is specified in terms of the north-south direction (Y) and the east-west direction (X). Two aspects of the table should be noted. First, the table shows that the number of points varies for each day, and this will affect our method of summarizing the location for a particular day. Second, the noted locations are dispersed over a fairly large geographic area. This dispersion could be due to movement of particular units within the battalion's area of operations, to movement of the entire battalion, or to the physical dispersion of men throughout the battalion's area of operation.

Several measures are available to summarize information on both the overall battalion location and the noted dispersion, and we have employed three to analyze the data in the table. First, the centroid is defined as the point having the arithmetic average of all available first coordinates as its first coordinate and the arithmetic average of all available second coordinates as its second coordinate . Several variations of this measure of centrality are possible, and all depend on taking weighted averages of coordinates rather than simple arithmetic averages. We could, for example, weight each coordinate according to the number of men in the unit for which the coordinate was reported. We applied some of these weighting schemes to the data but found that they made little difference in specifying a battalion's location. In addition, two measures of dispersion are presented in the table: the standard distance and the scatter. The standard distance is the square root of the sum of the variances of the first and second coordinates and may be viewed as the quadratic mean of the distances from the centroid to the individual locations. The scatter is the product of the standard deviations of the first and second coordinates.

These three statistics present a major problem in that consideration of optimal statistical properties (unbiasedness, efficiency, etc.) must be based on the assumption that the reported points constitute a random sample of all the points representing the location of men assigned to a battalion. We do not believe, however, that the data available for a given day constitute a random sample for several reasons, including differential reporting of locations for units within a battalion and differential availability of records for the ESG to abstract. Because of the nonrandom nature of the sample, we have concentrated on establishing a valid estimate of the centroid for the battalion area of operations on each day, de-emphasizing the measures of dispersion of the battalions.

4. Cluster Analysis

One effect of the nonrandom nature of the data demonstrated by Table 4 is the apparent rather than real movement of the battalion that occurs when successive daily centroids are compared. Examination of the daily centroids suggests that the battalion has undergone substantial movement during the days that are presented. A visual inspection of the individual coordinates, however, shows that this change in the centroids is due not to movement but rather to a difference in the number of times that the same locations have been reported on different days. Cluster analysis provides a method for reducing the effect of this differential sampling. We have employed this method on the coordinates for the given days and the resultant clusters are summarized in Table 5. A distance matrix is calculated for a particular day which contains the distance between all pairs of reported coordinates. The matrix is searched for the minimum distance, and when this distance is less than 10 km, the pair of points identified is replaced by their centroid. The process is repeated until the minimum of all the distances between two coordinates is less than 10 km. In this manner, we assign a location to a battalion on a particular day. The following table presents the results of applying this procedure to the data in table 4.

Table 5
Results from Clustering the Location Information for the First Battalion List on the AOP Master List 2/6/67 until 2/10/67.

8

Cluster:	2/06/67		2/07/67		2/08/67		2/09/67		2/10/67	
	X	Y	X	Y	X	Y	X	Y	X	Y
1	363.0	365.0	362.0	366.0	361.0	367.8	361.0	366.8	363.5	366.8
2	396.9	348.6	397.0	349.0	397.4	349.7	397.0	349.0	397.4	349.7
Centroid	380.0	356.8	379.5	357.5	379.2	358.8	379.0	357.9	380.5	358.3

Cluster analysis yields a series of daily locations for which the day to day movement is reduced substantially from that shown by centroids and therefore more nearly approximates the real rather than the apparent movements of the battalion. These algorithms will require further exploration and developement, but the process should change only minimally from that described.

B. Agent Orange Applications

The assignment of daily exposure scores to veterans involves calculating their distances from herbicide applications on the specified day. Each veteran will be assigned the location of his battalion for the day in question, and the process of assigning locations to battalions was described above. In this section we discuss the data available on herbicide applications and the methods for using these data to assign the daily exposure scores.

There are two main sources of information on herbicide applications in Vietnam which generally are referred to as the "Herbs Tape" and the "Services Herbs Tape." The Herbs Tape contains data on applications of herbicides by Operation Ranch Hand, the Air Force operation that applied herbicides in Vietnam using specially equipped fixed-wing aircraft. The data were compiled initially by the National Academy of Sciences, and there have been several attempts to validate the information on the tape. The latest validation studies were done in Australia and included a computer imaging of satellite photographs to analyze vegetation stress and its relationship to the data on the tape. These studies conclude that while the data appear to be consistent with the information available for validation, these sources are not sufficient to allow a definitive study. We plan no further attempts to validate these data.

We have tried to simplify the data structure employed on the Herbs Tape by creating a file of herbicide applications to consist of a set of records, each reporting the type of herbicide used, the date of spraying, the point at which the aircraft turned on the spray, and the location at which the spray was turned off. Each record of this type is referred to as an "herbicide application" and distances are calculated according to these records.

The Services Herbs Tape contains information on herbicide applications by specific units within Vietnam. These applications include helicopter or jeep based sprayings, leaks, and other ground sprayings. We created a file of herbicide applications for this tape also to correspond to the file created for the Herbs Tape. We have no way to judge the completeness of these data but ESG continually updates the file. Our present data were supplied by ESG in late 1983, and we expect to receive at least one more update before we complete the procedures for assigning exposure classifications.

We have combined the two files of herbicide applications to create a single file containing all of the data from both sources of information in order to simplify the algorithms for matching locations of individuals to locations of herbicide applications. Using the individual's assigned location on the day in question, we determine how far this is from all herbicide applications that occurred within 60 days and within 8 km of the noted location. Two different mathematical formulations, the "exponential-decay model" and the "hits model" will be used to assign the veteran a daily exposure score.

V. STATISTICAL MODELS

Ranking the likelihood of Agent Orange exposure involves linking data on battalion daily locations and the date and location of Agent Orange applications. To combine these data to estimate relative exposure likelihoods involves making assumptions about the influence of time and distance from the application on the probability of exposure. In other words, assumptions must be made about the half-life and geographic dispersal of Agent Orange.

A. Exponential Decay Models

For each day we have knowledge of a set of herbicide applications that contribute to that day's exposure score: they occurred within 60 days and 8 km of the the time and location under consideration. Denote these applications:

$$A(i), \quad i=1,\ldots,k$$

For each application, we calculate the distance from the battalion's location to that application. If the application has a single coordinate associated with it (as is the case for many entries from the Service Herbs Tape), the distance is calculated in the usual manner. If the application has two coordinates associated with it, the distance is calculated as the minimum distance from the noted location to the line segment resulting from connecting the two coordinates.

If a veteran is noted as being at a distance d(i) from the application and the application occurred at time t(i) in the past 60 days, then the most general form of the exponential-decay model is as follows:

$$E(i) = \exp[\, a + b*d(i) + c*t(i) \,]$$

This model has a certain amount of face validity: Many biological variables exhibit an exponential decay with time (Reference), and Air Force calibration tests of applications of Agent Orange using aircraft like those used in Vietnam indicate that the amount of herbicide reaching the ground at a specified distance from the flight path of the aircraft is consistent with an exponential model (Ref).

The intercept "a" in the general model specifies the amount of "exposure" when both time and distance are zero. We have chosen to eliminate this parameter since we wish to establish the exposure rank of the veteran relative to other veterans in the study, and this parameter has no effect on the ranking of individual scores. In addition, this approach acknowledges the biologic reality that we do not know the actual exposure of men in battalions located directly below spraying missions.

We have chosen the following model to assign exposure scores in relation to each application of herbicides associated with a daily location:

$$E(i) = \exp[\, b*d(i) + c*t(i) \,]$$

This formulation assigns the veteran a score of 1 when the time and distance from an application is 0, and the score decreases as both parameters increase.

Time Parameter: We used two separate approaches to establish the range of reasonable coefficients for the time parameter and the distance parameter. For the time parameter we considered possible half-lives subject to the criteria that the values at time equal to 60 days be essentially zero. The chosen half-lives of from 1-10 days are consistent with available ecologic and toxicologic data (refs). Table 6 summarizes the estimated coefficients for various hypothesized half-lives and also presents the value specified by each model at 60 days.

Table 6
Coefficients Associated with the Time Parameter in Exponential-Decay Model For Specified Half-Lives

Half-Life (days)	Coefficient	Value at 60 days
1	-0.69314	.00000
2	-0.34657	.00000
3	-0.23104	.00000
4	-0.17328	0.00003
5	-0.13862	0.00024
6	-0.11552	0.00098
7	-0.09902	0.00263
8	-0.08664	0.00552
9	-0.07701	0.00984
10	-0.06931	0.01563

The condition that the value at 60 days be zero is met by half-lives of 10 days or less. We will discuss the implication of adopting the different parameter values further under section VI.B.2..

Distance Parameter: We used the data from the Air Force calibration study (Ref) to establish the magnitude of the parameter associated with distance in the exponential-decay model and also to establish the distance beyond which we would consider the exposure to be essentially zero. The actual coverage in gallons of herbicides per acre at specified distances from the aircraft flight line was presented in the Air Force report in graphic form for each calibration run. The number of gallons of herbicides per acre was estimated from the graphs at distances of 100, 200, 300, 400, and 500 meters, and these data were used to fit an exponential-decay model with the time parameter fixed as zero. The results are shown in Table 7.

Table 7
Summarization of Distance Parameters Estimated from the U.S. Air Force Calibration Tests.

Coefficient	R Square	Value at 8 km	Conditions
-8.74	0.877	.00000	Inwind
-1.92	0.797	.00000	Crosswind, 90 degrees
-3.34	0.987	.00000	Crosswind, 270 degrees
-2.27	0.741	.00000	Crosswind, 270 degrees
-1.76	0.769	.00000	Crosswind, 45 degrees
-3.08	0.924	.00000	Crosswind, 90 degrees
-1.20	0.494	.00007	Crosswind, 90 degrees

We found all of the parameter values to be essentially zero at 8 km, and all matching of troop locations to herbicide applications are restricted to this distance.

B. "Hits" Models

The second model, the hits model, specifies the following formula:

$$E(i) = \begin{cases} 1 \text{ if } d(i) \quad d \quad \text{and} \quad t(i) \quad t \\ 0 \text{ otherwise} \end{cases}$$

In other words, if a veteran's battalion passes within d kilometers of a recent Agent Orange spray mission path on a given day (within t days of the application), he will be assigned one exposure likelihood "point" for that day.

C. Ranking Veterans' Likelihood of Exposure

For both exponential-decay and hits models we summarize the numbers E(i) for all of the applications noted. We use the formula

$$E = E(1)+E(2)+...+E(k)$$

as the summary exposure score for each day the veteran was present in III Corps.

These daily scores for each veteran will then be totaled over his entire stay in Vietnam. We will then rank these cumulative scores and use the ranks as an index of the likelihood of Agent Orange exposure. The analysis will assess the association between these ranks and the risk of the various health outcomes under study. At least one exponential-decay model and one hits model will be used.

Several important details about the statistical models used to assess exposure are discussed in relation to misclassification below. Final specification of these details, including the selection of time and distance parameters for the models, will await completion of computer simulations estimating the effects of misclassification.

VI. MISCLASSIFICATION OF EXPOSURE

The issue of misclassification of Agent Orange exposure is central to the interpretation of results from the Agent Orange Study. There are two main reasons why accurate estimation of the "true", biologically effective exposure to Agent Orange or dioxin among Vietnam veterans is impossible. First, individual veterans cannot be precisely located in relation to Agent Orange applications. The military records were not collected for the purposes of epidemiologic research and allow only assigning veterans the "average" location of their battalions. Second, even if we knew the exact location of veterans in relation to every application of Agent Orange in Vietnam, we could not accurately estimate the actual exposure to Agent Orange or dioxin. Not enough is known of the ecologic, toxicologic, or physiologic properties of Agent Orange to know what constitutes a certain "exposure level."

Misclassification is of concern because it tends to obscure real associations of exposures and outcomes. This concern has influenced the design and implementation of the Agent Orange Study and is one major rationale for CDC's Vietnam Experience Study. It also is one reason we favor ranking veterans as to their "likelihood" or "opportunity" of Agent Orange exposure, rather than attempting to estimate veterans actual dioxin exposure in any absolute terms.

This section discusses some of the general effects of misclassification of Agent Orange Exposure, as well as several possible sources of misclassification. Particular attention is paid to estimating the likely magnitude of the misclassification.

A) General Effects of Misclassification

The statistical framework required to establish whether Agent Orange has produced detrimental effects in veterans will employ three random variables:

X1 = an approximate index of exposure
X2 = true exposure
X3 = an outcome of interest

X1 is the summarization of the mathematical model specifying the relationship between exposure and the time and distance from a particular application of herbicides. We never can know the relationship between X2 and X3, but we will draw inferences about this relationship solely on the basis of the relationship between X1 and X3. We also never can know the relationship between X1 and X2, but we may examine the effect of this relationship by making certain assumptions. For example, let us assume that the distribution of these variables is multivariate normal and that there is no relationship between the approximate index of exposure and the outcome if we control for true exposure. Under these assumptions, the partial correlation between X1 and X3 holding X2 constant is zero. A simple algebraic solution using the definition of partial correlation yields the following formula:

$$corr(X1,X3) = corr(X1,X2) * corr(X2,X3)$$

From this formula we see that the true relationship between exposure and outcome will exceed that observed in our study. Moreover, for any observed magnitude of an exposure-outcome association, the "true" magnitude of association can be estimated using the formula, assuming various correlations between the true and estimated exposures. While this procedure will not be used to "adjust" the relative risks or other measures of association in our study reports, it will be discussed in the general sense to emphasize the likely presence of misclassification and the potential magnitude of its effect.

A related issue that is currently being examined is the influence of misclassification on the relative statistical power of categorizing Agent Orange exposure likelihood rankings in fairly broad groups, such as halves or quartiles, as opposed to relying on the ranks themselves as a non-parametric ordinal measure of exposure.

VI. B. Sources of Misclassification

1. Statistical Models

Since we do not know precisely what constitutes exposure to Agent Orange or how this exposure should be measured, we probably never will be able to determine the definitive model. We must, however, examine the extent to which an incorrect model specification affects the classification of individuals. We will design a simulation study involving this type of misclassification when we obtain sufficient data to specify the statistical distribution of time and distance from herbicide applications for the study participants. The general approach will be to simulate an underlying exposure model and to classify exposure using a series of models known to be incorrect.

For example, if we suppose that the true relationship between exposure and time and distance is expressed by the following linear model:

$$E = a + b*d + c*t + e$$

where E = exposure
d = distance
t = time
e = error

and we use an exponential-decay model to classify individuals, this incorrect model specification will introduce misclassification. This type of misclassification is one reason for our decision to use either ranks or broad exposure categories of ranks for analysis since we believe a non-parametric index of exposure likelihood will not be substantially affected by this type of error. Our simulation studies will evaluate this belief.

VI. B. 2. Incorrect Time and Distance Parameters

Even if we select the correct model for classifying exposure, we face the possibility that the parameters we employ in the model are incorrect. We also will explore this possibility using a simulation study. Our initial design for this simulation study involves assigning daily exposure scores to battalions and summarizing these scores for the two years from 1-1-67 through 12-31-68. We use the average rather than the sum of the daily scores, since the number of days for which location information is available currently varies among the battalions (Table 1). This simulation allows us to examine whether the choice of parameters for the models used for assigning exposure changes the exposure classification significantly. Table 8 summarizes some results for possible sets of coefficients for the exponential-decay model. In the table we note only whether a battalion is in the upper or lower half of the summary scores.

The selection of this particular set of parameters for the model was discussed earlier in this report. The rankings into high or low categories did not change appreciably with different sets of coefficients. Further simulations will assess the influence of more extreme coefficients.

In addition to placing battalions in the upper and lower half of the ranking of summary scores, the actual ranks were used to calculate a rank correlation matrix which is shown in Table 9. The correlation matrix also verifies that the rankings are consistent over a broad range of coefficients.

Table 8
Results of Ranking Battalions Using Specified Values of the Parameters for the Exponential-Decay Model.

Parameters:																					
Time	1	1	1	1	1	1	1	2	2	2	2	2	2	2	3	3	3	3	3	3	3
Dist.	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
Model No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Bat:																					
1	H	H	H	H	H	H	H	H	H	H	H	H	H	H	L	H	H	H	H	H	L
2	L	L	H	L	L	H	L	H	L	L	L	L	L	L	H	L	L	L	L	L	L
3	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H
4	H	L	L	L	L	L	L	H	H	H	H	L	H	L	H	L	H	H	L	H	L
5	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
6	L	H	L	H	H	L	H	L	L	L	L	L	L	L	L	L	L	L	L	L	H
7	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H
8	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H
9	L	L	L	L	L	L	L	L	L	L	L	H	L	H	L	H	L	L	H	L	H
10	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
11	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
12	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
13	L	L	L	L	L	L	L	L	L	L	L	L	L	L	H	L	L	L	L	L	L
14	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
15	H	H	H	H	H	H	H	L	H	H	H	H	H	H	L	H	H	H	H	H	H
16	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H
18	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
20	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H
21	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L
24	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H
26	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H

Note: The following are the parameter values used:

Time (1) -0.69314
(2) -0.23104
(3) -0.06931

Distance: (1) -8.74
(2) -1.92
(3) -3.34
(4) -2.27
(5) -1.76
(6) -3.08
(7) -1.20

Table 9
Rank Correlation Matrix for the Models
Presented in Table 8
9
Model:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	1.0	0.9	1.0	0.9	0.9	1.0	0.9	1.0	0.9	1.0	1.0	0.9	1.0	0.9	0.8	0.9	1.0	1.0	0.9	1.0	0.8
2	0.9	1.0	1.0	1.0	1.0	1.0	1.0	0.9	1.0	0.9	1.0	1.0	0.9	1.0	0.7	0.9	0.9	0.9	0.9	0.9	0.9
3	1.0	1.0	1.0	1.0	1.0	1.0	0.9	0.9	1.0	1.0	1.0	0.9	1.0	0.9	0.8	0.9	0.9	0.9	0.9	1.0	0.8
4	0.9	1.0	1.0	1.0	1.0	1.0	1.0	0.9	1.0	0.9	1.0	1.0	0.9	0.9	0.7	0.9	0.9	0.9	0.9	0.9	0.8
5	0.9	1.0	1.0	1.0	1.0	1.0	1.0	0.8	1.0	0.9	1.0	1.0	0.9	1.0	0.7	0.9	0.9	0.9	0.9	0.9	0.9
6	1.0	1.0	1.0	1.0	1.0	1.0	0.9	0.9	1.0	1.0	1.0	0.9	1.0	0.9	0.8	0.9	0.9	0.9	0.9	0.9	0.8
7	0.9	1.0	0.9	1.0	1.0	0.9	1.0	0.8	0.9	0.9	0.9	1.0	0.9	1.0	0.6	0.9	0.9	0.9	0.9	0.9	0.9
8	1.0	0.9	0.9	0.9	0.8	0.9	0.8	1.0	0.9	0.9	0.9	0.8	0.9	0.8	0.9	0.9	1.0	0.9	0.8	1.0	0.8
9	0.9	1.0	1.0	1.0	1.0	1.0	0.9	0.9	1.0	1.0	1.0	1.0	1.0	1.0	0.7	1.0	0.9	1.0	1.0	0.9	0.9
10	1.0	0.9	1.0	0.9	0.9	1.0	0.9	0.9	1.0	1.0	1.0	0.9	1.0	0.9	0.8	1.0	1.0	1.0	1.0	1.0	0.9
11	1.0	1.0	1.0	1.0	1.0	1.0	0.9	0.9	1.0	1.0	1.0	1.0	1.0	1.0	0.7	1.0	1.0	1.0	1.0	1.0	0.9
12	0.9	1.0	0.9	1.0	1.0	0.9	1.0	0.8	1.0	0.9	1.0	1.0	1.0	1.0	0.7	1.0	0.9	1.0	1.0	0.9	1.0
13	1.0	0.9	1.0	0.9	0.9	1.0	0.9	0.9	1.0	1.0	1.0	1.0	1.0	0.9	0.8	1.0	1.0	1.0	1.0	1.0	0.9
14	0.9	1.0	0.9	0.9	1.0	0.9	1.0	0.8	1.0	0.9	1.0	1.0	0.9	1.0	0.6	1.0	0.9	1.0	1.0	0.9	1.0
15	0.8	0.7	0.8	0.7	0.7	0.8	0.6	0.9	0.7	0.8	0.7	0.7	0.8	0.6	1.0	0.7	0.9	0.8	0.7	0.9	0.6
16	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	1.0	1.0	1.0	1.0	1.0	1.0	0.7	1.0	1.0	1.0	1.0	1.0	1.0
17	1.0	0.9	0.9	0.9	0.9	0.9	0.9	1.0	0.9	1.0	1.0	0.9	1.0	0.9	0.9	1.0	1.0	1.0	0.9	1.0	0.9
18	1.0	0.9	0.9	0.9	0.9	0.9	0.9	0.9	1.0	1.0	1.0	1.0	1.0	1.0	0.8	1.0	1.0	1.0	1.0	1.0	0.9
19	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.8	1.0	1.0	1.0	1.0	1.0	1.0	0.7	1.0	0.9	1.0	1.0	0.9	1.0
20	1.0	0.9	1.0	0.9	0.9	0.9	0.9	1.0	0.9	1.0	1.0	0.9	1.0	0.9	0.9	1.0	1.0	1.0	0.9	1.0	0.9
21	0.8	0.9	0.8	0.8	0.9	0.8	0.9	0.8	0.9	0.9	0.9	1.0	0.9	1.0	0.6	1.0	0.9	0.9	1.0	0.9	1.0

6

The following table presents results for ranking the battalions using the hits model with several different sets of parameters. As with the decay model the ranking into the top and bottom halves was consistent over a broad range of parameter choices.

Table 10
Results of Ranking Battalions Using Several Different Versions of the Hits Model.

TIME	1	1	1	2	2	2	3	3	3
DIST	1	2	3	1	2	3	1	2	3
BATTALION NUMBER:	/----------				RANK				----------/
1 -	L*	H	H	L	H	H	H	H	H
2 -	L*	L	H	H	L	H	H	H	H
3 -	H	L	H	H	H	H	L	H	H
4 -	L*	L	L	L	L	L	L	L	L
5 -	L*	H	H	L	H	H	L	L	L
6 -	L*	H	H	L	H	H	L	H	H
7 -	H	H	L	H	L	L	L	L	H
8 -	H	L	L	H	H	H	H	H	L
9 -	L*	H	H	L	H	H	H	H	H
10 -	L*	L	L	H	L	L	H	L	L
11 -	L*	L	L	L	L	L	L	L	L
12 -	L*	L	L	L	L	L	L	L	L
13 -	L*	L	L	L	L	L	L	L	L
14 -	L*	L	L	L	L	L	L	L	L
15 -	H	H	H	H	H	H	H	H	H
16 -	H	H	H	H	L	L	H	H	H
18 -	H*	L	L	L	L	L	L	L	L
20 -	H	H	H	H	H	H	H	H	H
21 -	H*	L	L	L	L	L	L	L	L
24 -	H	H	H	H	H	H	H	H	H
26 -	H	H	L	H	H	L	H	L	L

Note: * = tied rank
Parameters used in the specified models

TIME (1) 1 day
(2) 3 day
(3) 10 day

DISTANCE (1) 2 km
(2) 5 km
(3) 7 km

The results for the exponential-decay model and the hits model presented above may not hold for the ranking of individuals since the distribution of time and distances from herbicide applications for individuals probably will not resemble that for the battalions. We plan a further examination of the simulation studies when additional data on individuals' movements among units allow for more realistic models.

VI. B. 3. Battalion Location Data

In addition to the misclassifications due to incorrect models or incorrect coefficients used in the model, a major source of misclassification results from specifying locations inaccurately or imprecisely. When an individual's location is specified incorrectly or the location of an herbicide application is specified incorrectly or both, the result is an error in calculating the distance between the individual's location and the location of the herbicide application. The easiest way to study the joint effect of these two types of error is to simulate the error in distance calculation. This process in being designed.

VII. CONCLUSIONS

This report described several specific changes in procedures for the selection of participants in the Agent Orange Study. In addition, it provided an update of other information related to exposure assessment, including the results of several pretests of methods and our plans for further evaluations of these issues. We welcome advice and comments.

FIGURE 1
AGENT ORANGE STUDY
OVERVIEW OF MODIFIED METHODS