EDGEWOOD

CHEMICAL BIOLOGICAL CENTER

U.S. ARMY SOLDIER AND BIOLOGICAL CHEMICAL COMMAND

ECBC-TR-337

MATHEMATICAL LIMITS ON DIFFERENCES BETWEEN A POPULATION AND A SUBPOPULATION

Ronald B. Crosier

RESEARCH AND TECHNOLOGY DIRECTORATE

20031001 203

August 2003

Approved for public release;
distribution is unlimited.

Aberdeen Proving Ground, MD 21010-5424

Disclaimer

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave Blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	2003 August	Final; 01 Sep – 02 Apr

4. TITLE AND SUBTITLE
Mathematical Limits on Differences Between a Population and a Subpopulation

5. FUNDING NUMBERS
C-DE-AI01-00NN20105,
PR-01-01NN20105.001

6. AUTHOR(S)
Crosier, Ronald B.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
DIR, ECBC, ATTN: AMSSB-RRT-DP, APG, MD 21010-5424

8. PERFORMING ORGANIZATION REPORT NUMBER
ECBC-TR-337

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
DOE, ATTN: NNSA NA-22 (Dr. Terry Creque), 1000 Independence Avenue, SW, Washington DC 20585

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
DOE TPOC: Sandia Corporation dba Sandia National Laboratories, ATTN: John E. Brockmann, Mail Stop 0836, PO Box 5800, Albuquerque, NM 87185-0836

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
Civil defense planning requires estimates of the toxicity of chemical warfare agents to the general public, but the current toxicity estimates are for male soldiers. Individual susceptibilities for both the general population and the military subpopulation are modeled by a lognormal distribution. Historical military demographics are used to estimate the size of the subpopulation from which military personnel are drawn. Limits on the median effective dose and the probit slope of a subpopulation are determined as a function of the subpopulation size. The method of converting toxicity estimates is not limited to chemical warfare agents. Further, it may be used to convert toxicity estimates for the general population to toxicity estimates for a subpopulation; for example, the elderly. The modeling of subpopulations is not limited to toxicological applications. Blood lead data from the second National Health and Nutrition Examination Survey are compared to the subpopulation model.

14. SUBJECT TERMS
National Health and Nutrition Examination Survey (NHANES)

Lognormal distribution	Risk assessment	Bell model
Normal distribution	Demographics	Blood lead
Conversion factor	Subpopulation	Toxicity
Chemical warfare	Probit slope	

15. NUMBER OF PAGES
24

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	**UNCLASSIFIED**	**UNCLASSIFIED**	**UL**

Blank

PREFACE

The work described in this report was authorized under Intra-agency Agreement Contract No. DE-AI01-00NN20105 (Project No. 01-01NN20105.001) with the U.S. Department of Energy, National Nuclear Security Administration, Chemical and Biological National Security Program. The work was started September 2001 and completed in April 2002.

The use of either trade or manufacturers' names in this report does not constitute an official endorsement of any commercial products. This report may not be cited for purposes of advertisement.

This report has been approved for public release. Registered users should request additional copies from the Defense Technical Information Center; unregistered users should direct such requests to the National Technical Information Service.

Acknowledgments

The author thanks Dr. Sharon A. Reutter, and Douglas R. Sommerville, Research and Technology Directorate, U.S. Army Edgewood Chemical Biological Center, for technical advice and assistance.

Blank

CONTENTS

FIGURES

TABLE

MATHEMATICAL LIMITS ON DIFFERENCES BETWEEN A POPULATION AND A SUBPOPULATION

1. INTRODUCTION

The U.S. Department of Energy's Chemical-Biological Nonproliferation Program has the task of improving the U.S. capability to prepare for and respond to the use of chemical and biological warfare agents against the civilian population. The Modeling Subgroup of the Technology Development Program Area is responsible for developing models of atmospheric transport and dispersion of chemical warfare agents. To assess casualties, these models require estimates of chemical warfare agent toxicity to the civilian population. However, the present chemical warfare agent toxicity estimates (Grotte and Yang 2001) are for male soldiers. In the absence of data relevant to the required soldier-to-civilian adjustment, either no adjustment is done or a guess is made. Concern that guesses at the required adjustment may exceed what is mathematically possible led to this report. Because a subpopulation is part of the population, there are mathematical limits on how much a subpopulation can differ from the population. Suppose, for example, that a resistant subpopulation is 20% of the entire population. The extreme placement for this subpopulation would be that it is the upper 20% of the population. Then the median of this subpopulation would be at the 90^{th} percentile of the population. For a normal distribution, the 90^{th} percentile is 1.28 standard deviations above the mean. For any distribution with finite variance, 90% of the population must lie within 3.16 standard deviations of the mean (by Chebyshev's Inequality; see, for example, Mood, Graybill, and Boes 1974); hence, the 90^{th} percentile of the population (and the median of the subpopulation) cannot be farther away from the mean than 3.16 standard deviations. Although the placement of a subpopulation into the tail of the population distribution is not realistic, it establishes the existence of limits on the difference between a population and a subpopulation. Further, it reveals three relevant factors: the size of the subpopulation, the standard deviation of the population, and the distribution of the population. The next section develops the theory and notation for a more reasonable model.

2. THEORY AND NOTATION

The susceptibility of the population to a toxicant was modeled by a lognormal distribution of effective doses. (The theory and methods are also applicable to dosages.) Toxicologists characterize a lognormal distribution by its median effective dose (ED_{50}) and its probit slope, m. The probit slope is the reciprocal of the standard deviation of log (effective dose), where log is the common (base 10) logarithm. Individual susceptibilities are given in Z units of the population by

$$Z = m_{pop}\,[\log(ED) - \log(ED_{50})], \tag{1}$$

where m_{pop} is the probit slope of the population and *ED* is the effective dose for an individual. Thus, the population is represented by a standard (mean zero, variance one) normal distribution.

2.1 Subpopulation Model.

The distribution of effective doses for a subpopulation was modeled as a lognormal distribution—hence, the distribution of log (*ED*) for the subpopulation follows the bell-shaped normal curve. Let μ and σ be the subpopulation mean and the subpopulation standard deviation in *Z* units of the population—that is, μ and σ are calculated from the effective doses of the subpopulation after transforming the effective doses by (1). Thus,

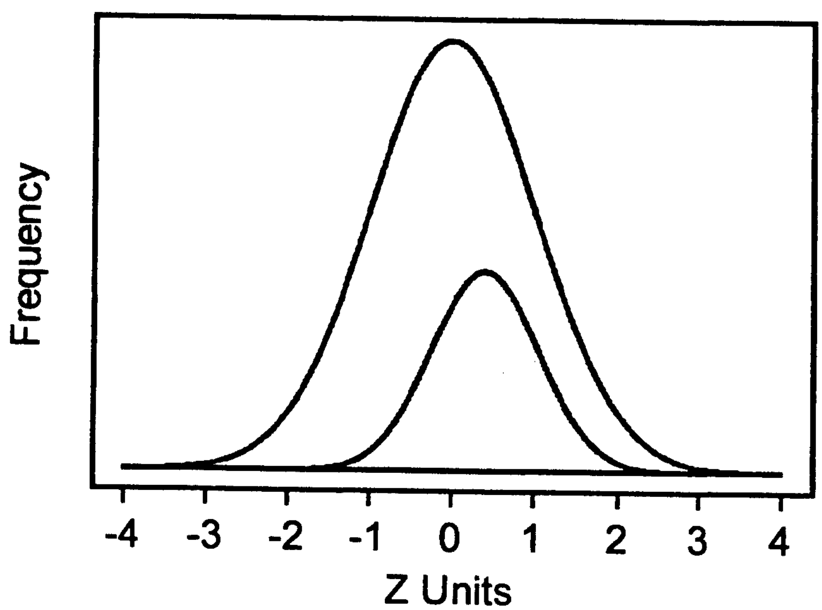
$$\mu = m_{pop}\,[\log(\text{subpopulation } ED_{50}) - \log(\text{population } ED_{50})] \tag{2}$$

and

$$\sigma = \sigma_{sub} / \sigma_{pop} = m_{pop} / m_{sub} \tag{3}$$

where σ_{sub} and σ_{pop} are the standard deviations of the subpopulation and the population, respectively, in log(*ED*) units and m_{sub} is the probit slope of the subpopulation. The size of the subpopulation, θ, is defined as a fraction of the population. Figure 1 shows a subpopulation of size θ = 0.3. The curves in Figure 1 are not probability densities but frequencies—normal curves fit to histograms—as described, for example, in chapter 5 of Dixon and Massey (1969).

Figure 1. Model for a Subpopulation of Size θ = 0.3



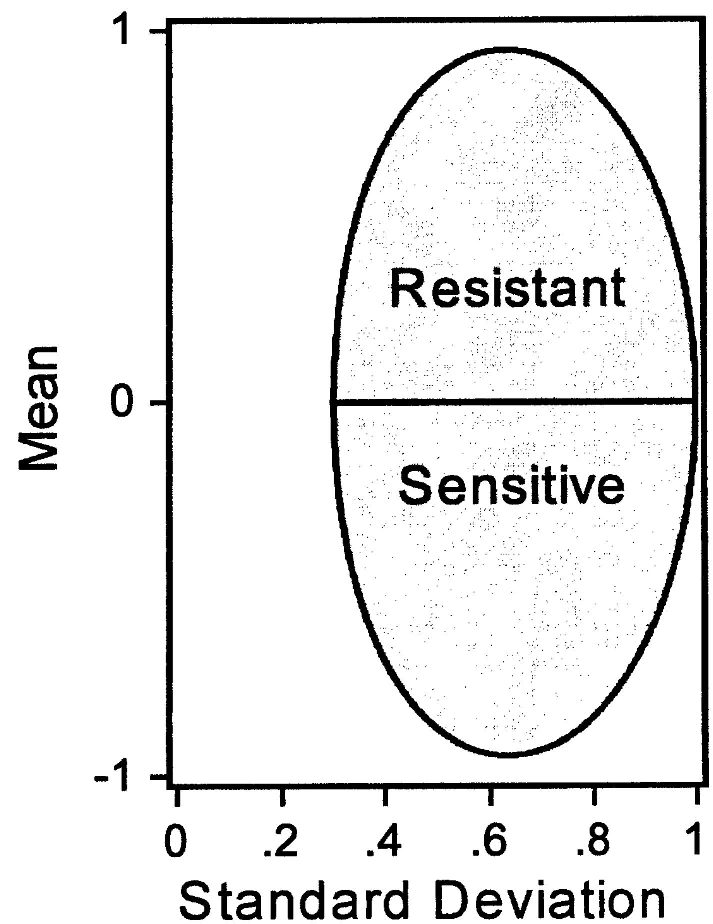
2.2 Feasible Values for Subpopulation Parameters.

The combinations of μ and σ that allow the subpopulation bell curve to remain entirely within (or under) the population bell curve constitute the feasible region for μ and σ. Figure 2 shows the feasible region (shaded) for a subpopulation of size $\theta = 0.3$. Crosier and Sommerville (2002) determined feasible values of μ and σ by numerical searches; Figure 2, however, was obtained from

$$\mu = \pm [-2(1-\sigma^2)\ln(\theta/\sigma)]^{1/2}, \quad (4)$$

where ln is the natural (base *e*) logarithm. Equation (4) gives the upper and lower limit of the feasible range for μ as a function of θ and σ. Appendix A gives the derivation of (4); the ranges of θ and σ are $0 < \theta \le \sigma < 1$. When μ is plotted on the y-axis, as in Figure 2, the feasible region is symmetrical about the x-axis; μ is positive for a resistant subpopulation and negative for a sensitive subpopulation. Henceforth, the term feasible region will be limited to either the resistant subpopulation case or the sensitive subpopulation case. It is only necessary to study one case; the results apply to the

Figure 2. Feasible Region of σ and μ for a Subpopulation of Size $\theta = 0.3$



other case by symmetry. Figures 3 (minimum σ), 4 (maximum σ), and 5 (maximum μ) show subpopulations that correspond to limits of the feasible region for μ and σ when $\theta = 0.3$. The case shown in Figure 3 (minimum σ) is highly unlikely. To obtain such a subpopulation, there would have to be a strong selection bias for individuals with effective doses near the population ED_{50}.

Note that Figure 4 has $\mu = 0$ and $\sigma = 1$; these are the expected values of the subpopulation parameters when the subpopulation is a random sample of the population. If one considers the age, sex, health, and physical fitness status of military personnel as irrelevant to their susceptibility to chemical warfare agents, then military personnel can be regarded as a random sample of the general population.

Figure 3. Subpopulation of Size $\theta = 0.3$ with Minimum Standard Deviation

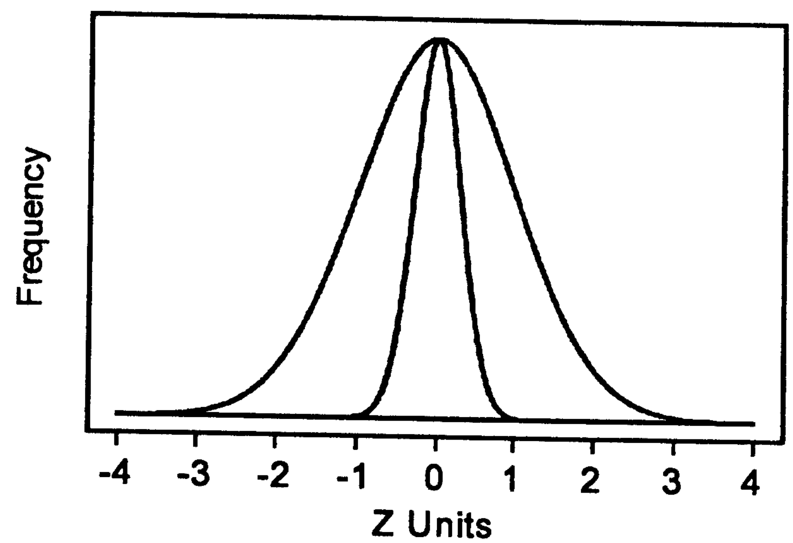


Figure 4. Subpopulation of Size $\theta = 0.3$ with Maximum Standard Deviation

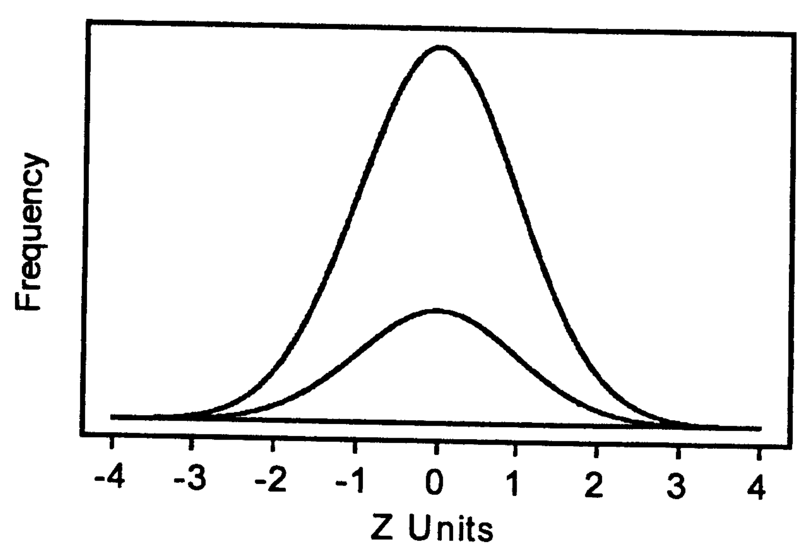
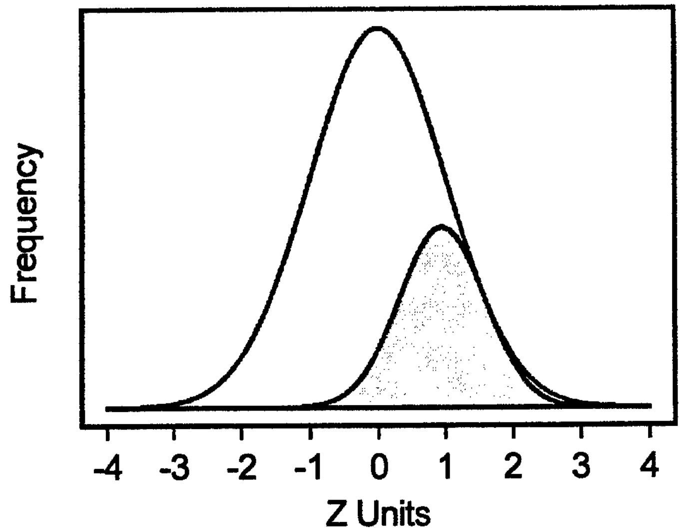


Figure 5. Subpopulation of Size $\theta = 0.3$ with Maximum Mean



2.3 Selection of Subpopulation Parameters.

Given the feasible region for a resistant or sensitive subpopulation of size θ, how does one select μ and σ to represent the subpopulation? There are many choices, which can be categorized under two philosophies: the worst-case combination of μ and σ, or typical values of μ and σ.

For conversion of a population ED_{50} to a resistant subpopulation ED_{50}, the largest conversion factor is obtained by selecting the value of σ—call it σ_x— that yields the maximum value for μ, denoted μ_x. Appendix B gives the mathematical derivation and numerical method used to find σ_x. For conversion of a resistant subpopulation ED_{50} to the population ED_{50}, the conversion factor will be less than one; the smallest conversion factor is obtained by selecting the value of σ that maximizes the ratio μ / σ. Denote the values of μ and σ of the largest ratio by μ_r and σ_r. Generation of μ_r and σ_r is discussed in Appendix C.

Two estimators of typical values for μ and σ are the mid-range and the centroid. The mid-range estimates are: $\mu_m = \mu_x / 2$ and $\sigma_m = (\theta + 1) / 2$. The formulas for the centroid estimates μ_c and σ_c are in the form of an integral divided by the area of the feasible region, which is also expressed as an integral. The integral for the numerator of μ_c has analytic solution $(9\,\theta - 6\ln(\theta) - \theta^3 - 8) / 9$, but the integrals for the numerator of σ_c and for the area of the feasible region were evaluated numerically.

Figure 1 is based on the centroid values $\mu_c = 0.403$ and $\sigma_c = 0.645$ for a resistant subpopulation of size $\theta = 0.3$, whereas Figure 5 uses the maximum-mean values $\mu_x = 0.946$ and $\sigma_x = 0.633$ for a resistant subpopulation of size $\theta = 0.3$. The maximum-mean case in Figure 5 might be appropriate for describing the physical fitness of the military personnel (because physical fitness is a selection criterion or requirement for military personnel), but it may be excessive for describing susceptibility to chemical warfare agents.

3. RESULTS

3.1 Subpopulation Parameters.

The table gives the standard deviations and means for resistant subpopulations; for sensitive subpopulations, multiply the means by –1. The units for the standard deviations and means in the table are Z units of the population.

3.2 Conversion of Toxicity Estimates.

Suppose a population has an ED_{50} of 100 and a probit slope of 5. To convert these estimates to a sensitive subpopulation of size $\theta = 0.2$, I use the centroid values from the table, $\mu_c = -0.508$ and $\sigma_c = 0.591$, which are in Z units of the population. Equation (2) may be used to convert $\mu_c = -0.508$ to the median effective dose for the subpopulation. Substituting into equation (2) yields $-0.508 = 5$ [log(subpopulation ED_{50}) – log(100)], or subpopulation ED_{50} = antilog [$-0.508 / 5$ + log(100)] = 79 (rounded). Alternatively, one may develop a conversion factor to convert the population ED_{50} to the subpopulation ED_{50}. The factor to convert from effective dose A to effective dose B may be obtained from ED_A and ED_B expressed in Z units of the population—denoted Z_A and Z_B, respectively—by:

$$\text{Conversion Factor} = ED_B / ED_A = \text{antilog}\,[\,(Z_B - Z_A) / m_{pop}\,]. \qquad (5)$$

Applying (5) with $Z_A = 0$, $Z_B = -0.508$, and $m_{pop} = 5$ yields 0.791, so the ED_{50} of the subpopulation is (100)(0.791) = 79 (rounded). To convert the population probit slope to a probit slope for the subpopulation, I substitute $\sigma_c = 0.591$ from the table into equation (3): $0.591 = \sigma_{sub} / \sigma_{pop} = m_{pop} / m_{sub}$, or $m_{sub} = m_{pop} / 0.591$. Thus, dividing the population probit slope ($m_{pop} = 5$) by $\sigma_c = 0.591$ yields the subpopulation probit slope: $m_{sub} = 5 / 0.591 = 8$ (rounded).

Conversions from a subpopulation to the population require that the probit slope be converted before conversion of the ED_{50} because the population probit slope is needed to convert the ED_{50}. For example, suppose a resistant subpopulation of size $\theta = 0.3$ has an ED_{50} of 200 and a probit slope of 10. First convert the probit slope: from the table, $\sigma_c = 0.645$, and applying $m_{pop} = (m_{sub})(\sigma_c)$ gives $m_{pop} = (10)(0.645) = 6.45$. Then use $m_{pop} = 6.45$ to convert the ED_{50}: applying (5) with $Z_A = 0.403$, $Z_B = 0$, and $m_{pop} = 6.45$ yields a conversion factor of 0.866. Thus, the population has an ED_{50} of (200)(0.866) = 173 and a probit slope of 6 (rounded).

Table. Subpopulation Parameters

Size (θ)	Subpopulation Standard Deviations: Max Mean (σ_x)	Max Ratio (σ_r)	Centroid (σ_c)	Resistant Subpopulation Means[a]: Max Mean (μ_x)	Max Ratio (μ_r)	Centroid (μ_c)
0.001	0.285	0.002	0.446	3.223	1.000	1.408
0.002	0.301	0.003	0.450	3.020	1.000	1.318
0.003	0.312	0.005	0.452	2.896	1.000	1.263
0.004	0.320	0.007	0.455	2.805	1.000	1.222
0.005	0.327	0.008	0.456	2.733	1.000	1.190
0.006	0.333	0.010	0.458	2.672	1.000	1.164
0.007	0.338	0.012	0.460	2.621	1.000	1.141
0.008	0.343	0.013	0.461	2.575	1.000	1.120
0.009	0.347	0.015	0.463	2.535	1.000	1.102
0.010	0.351	0.016	0.464	2.498	1.000	1.086
0.020	0.381	0.033	0.475	2.245	0.999	0.973
0.030	0.402	0.049	0.484	2.086	0.998	0.902
0.040	0.419	0.066	0.492	1.968	0.996	0.850
0.050	0.434	0.082	0.500	1.873	0.993	0.807
0.060	0.447	0.098	0.507	1.793	0.990	0.772
0.070	0.458	0.115	0.514	1.723	0.987	0.741
0.080	0.469	0.131	0.520	1.661	0.983	0.714
0.090	0.480	0.147	0.527	1.605	0.978	0.689
0.100	0.489	0.163	0.533	1.554	0.974	0.667
0.110	0.499	0.178	0.539	1.507	0.968	0.646
0.120	0.507	0.194	0.545	1.463	0.962	0.627
0.130	0.516	0.210	0.551	1.422	0.956	0.609
0.140	0.524	0.225	0.557	1.384	0.949	0.592
0.150	0.532	0.240	0.563	1.347	0.942	0.576
0.160	0.540	0.255	0.569	1.313	0.935	0.561
0.170	0.547	0.270	0.575	1.280	0.927	0.547
0.180	0.555	0.285	0.580	1.248	0.919	0.533
0.190	0.562	0.300	0.586	1.218	0.910	0.520
0.200	0.569	0.314	0.591	1.189	0.901	0.508
0.250	0.602	0.383	0.619	1.059	0.853	0.451
0.300	0.633	0.447	0.645	0.946	0.800	0.403
0.350	0.663	0.507	0.672	0.846	0.743	0.360
0.400	0.691	0.563	0.698	0.756	0.683	0.321
0.450	0.719	0.614	0.723	0.673	0.623	0.286
0.500	0.746	0.662	0.749	0.596	0.562	0.253
0.550	0.772	0.707	0.774	0.523	0.501	0.222
0.600	0.798	0.748	0.799	0.455	0.441	0.193
0.650	0.824	0.787	0.825	0.390	0.381	0.166
0.700	0.849	0.823	0.850	0.328	0.323	0.139
0.750	0.875	0.857	0.875	0.269	0.266	0.114
0.800	0.900	0.889	0.900	0.212	0.210	0.090
0.850	0.925	0.919	0.925	0.156	0.156	0.066
0.900	0.950	0.947	0.950	0.103	0.103	0.044
0.950	0.975	0.974	0.975	0.051	0.051	0.021

[a] For sensitive subpopulations, multiply the means by −1.

3.3 Comparison to Blood Lead Data.

Blood lead data from the second National Health And Nutrition Examination Survey were analyzed in terms of subpopulation parameters and compared to the subpopulation model. A report by the National Center for Health Statistics, Annest, and Mahaffey (1984) gives geometric mean blood lead levels, geometric standard deviations, and population size for the target population (non-institutionalized civilians aged 6 months to 74 years) and for 134 subpopulations defined by one or more of the factors race, age, sex, income, and type of residence (central city, urban, rural). The geometric standard deviations were adjusted for analytical error before calculation of the subpopulation statistics (an estimate of analytical error was given in Appendix II of the report). Although blood lead levels for a homogeneous group follow a lognormal distribution (Hasselblad, Stead, and Galke 1980), there is no reason to believe that blood lead levels for the target population—or for the subpopulations defined by the factors above—will do so. However, comparing the subpopulation model to data that meet the model assumptions would seem pointless—such data cannot violate the derived mathematical limits. Figure 6 plots the absolute deviations of the subpopulation means from the population mean versus the subpopulation size. For comparison to the subpopulation model, the predicted maximum mean (μ_x) is plotted as a triangle, and the centroid estimate of the mean (μ_c) is plotted as a circle, in Figure 6. Figure 7 plots the subpopulation standard deviations versus subpopulation size. There is a trend

Figure 6. Subpopulation Mean Versus Subpopulation Size

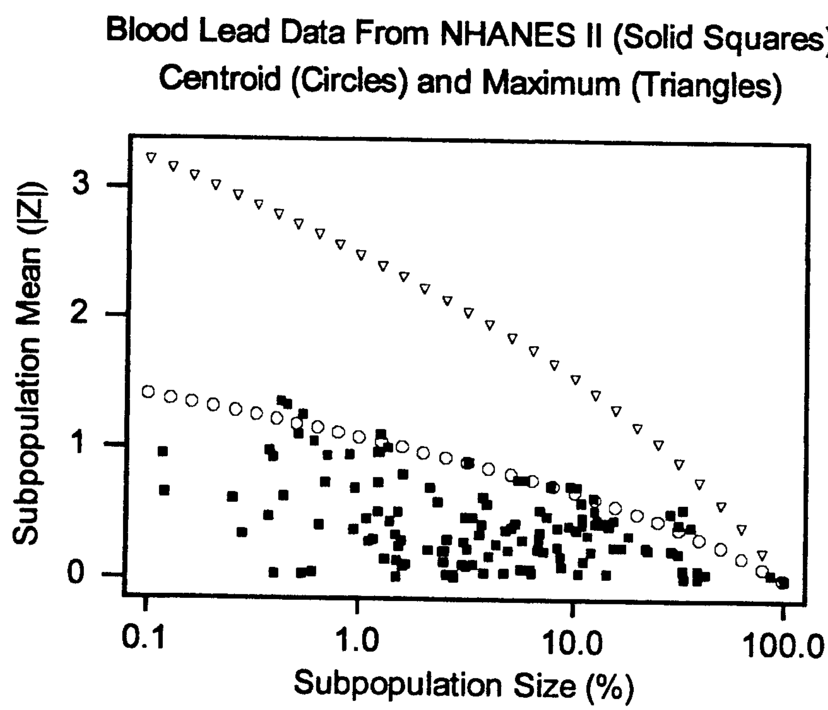
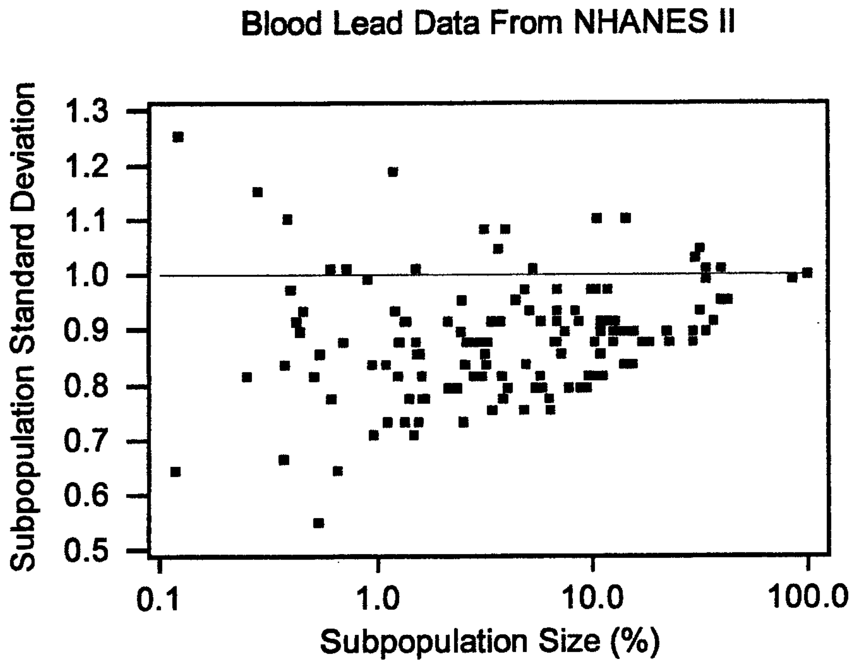


Figure 7. Subpopulation Standard Deviation Versus Subpopulation Size



toward smaller standard deviations for smaller subpopulations, but the subpopulation standard deviations are far from their lower bounds. The subpopulation model limits the range of σ to $\theta \leq \sigma < 1$, but 17 subpopulations have estimated standard deviations greater than one. Of these 17 subpopulations, nine have 95%, two-sided confidence intervals for σ that overlap one. Thus, the standard deviations greater than one appear to be partly due to random variation and partly due to the failure of the normality assumption (for logarithms of the data).

4. DISCUSSION

4.1 Mixed Distributions and Gender.

The subpopulation model uses a lognormal distribution for the susceptibility of individuals in the population. Although the model includes the case $\theta = 0.5$, the model is not intended to represent sex differences. Sex differences may be better modeled as two lognormal distributions—one for each sex. When combined, the two lognormal distributions may produce a mixed distribution rather than a lognormal distribution. A mixed distribution due to gender effects can be analyzed by applying the subpopulation model to each gender separately. The subpopulation model is intended to represent subpopulations created by selection (non-random sampling); it is not intended to represent subpopulations that have a known biological difference from the rest of the population.

4.2 Subpopulation Size Estimation.

For demographically defined subpopulations, estimation of subpopulation size is straightforward—for example, individuals aged 65 years or older constituted 12.4% of the U.S. population in 2000 (U.S. Census Bureau 2001a). However, this simple method is not appropriate for quantifying the size of the military subpopulation. Currently, the percentage of the population serving in the military is small; however, I consider individuals in the military to be randomly selected from a subpopulation of healthy, physically fit, young adults. Therefore, an estimate of the size of this subpopulation is needed. To obtain such an estimate Crosier and Sommerville (2002) examined historical military demographics.

U.S. military strength reached its peak of 12.1 million in 1945 (Dunnigan and Nofi 1994). The population of the United States in 1945 was 140 million (U.S. Census Bureau 2000). Therefore, about 8.6% of the U.S. population in 1945 was in the military. However, military personnel in World War II were nearly all men, so about 17% of the male population was in the military in 1945. Besides the men in the World War II military, there were other men who were qualified for military service but did not serve. Thus, $\theta = 0.17$ is a lower bound for the size of the subpopulation from which military personnel are selected.

U.S. men from 18 to 45 years old were liable for service in World War II (*Selective Service* 2001). This age group comprised 42% of the male population in 1999 (US Census Bureau 2001b). However, not every man in this age range is fit for military service. Thus, $\theta = 0.42$ is an upper bound. A reasonable estimate would be $\theta = (0.17 + 0.42)/2$, or $\theta = 0.30$. Because resistance to chemical warfare agents is not necessarily a function of physical size and strength, female soldiers can be regarded as randomly drawn from a resistant female subpopulation that is 30% of the total female population. The estimate $\theta = 0.30$ applies to the subpopulation from which military personnel are drawn. There are other resistant subpopulations, such as the working population. In 2000, the U.S. workforce was 49% of U.S. population (U.S. Census Bureau 2001a, 2001c). As in the case of the armed forces, the actual size of the workforce is an underestimate because many individuals who are capable of working are not in the workforce.

For toxicologists, the laboratory animal is a subpopulation of interest. An animal's health and susceptibility to toxicants varies over the animal's lifetime. The use of young, adult animals in toxicological studies creates a selection bias in the results that is unrelated to animal-to-human scaling. To quantify the magnitude of this selection bias, note that 14% of U.S. population was in the age range 15-24 years, inclusive, in 2000 (U.S. Census Bureau 2001a). The age range of animals in a toxicological study may be very narrow, but the animals are similar to other young adult animals, so $\theta = 0.15$ is a reasonable estimate for the size of the subpopulation of young, adult laboratory animals.

LITERATURE CITED

Crosier, R. B., and Sommerville, D. R., 2002: *Relationship Between Toxicity Values for the Military Population and Toxicity Values for the General Population.* Report ECBC-TR-224, Edgewood Chemical Biological Center. Aberdeen Proving Ground, Maryland.

Dixon, W. J., and Massey, F. J., Jr., 1969: *Introduction to Statistical Analysis,* 3rd edition. McGraw-Hill, New York.

Dunnigan, J. F., and Nofi, A. A., 1994: *Dirty Little Secrets of World War II.* William Morrow & Co., New York.

Grotte, J. H., and Yang, L. I., Eds., 2001: *Report of the Workshop on Chemical Agent Toxicity for Acute Effects, Institute for Defense Analyses, May 11-12, 1998.* Institute of Defense Analyses Document # D2176. Institute for Defense Analyses, Alexandria, VA.

Hasselblad, V., Stead, A.G., and Galke, W., 1980: Analysis of Coarsely Grouped Data From the Lognormal Distribution. *Journal of the American Statistical Association* 75, 771-778.

Mood, A.M., Graybill, F.A., and Boes, D.C., 1974: *Introduction to the Theory of Statistics*, 3rd edition. McGraw-Hill, New York.

National Center for Health Statistics, Annest, J.L., and Mahaffey, K., 1984: Blood lead levels for persons ages 6 months—74 years, United States, 1976-1980. *Vital and Health Statistics.* Series 11, No. 233. DHHS Pub. No. (PHS) 84-1683. Public Health Service. U.S. Government Printing Office, Washington, D.C.

Selective Service, 2001: The Columbia Encyclopedia, 6th edition. Columbia University Press, New York.

U.S. Census Bureau, 2000: *Historical National Population Estimates: July 1, 1900 to July 1, 1999*

U.S. Census Bureau, 2001a: *Table DP-1. Profile of General Demographic Characteristics for the United States: 2000.*

U.S. Census Bureau, 2001b: Table No. 13, Resident Population by Sex and Age: 1999. In *Statistical Abstract of the United States: 2000.*

U.S. Census Bureau, 2001c: *Table QT-03. Profile of Selected Economic Characteristics: 2000.*

Blank

APPENDIX A

FEASIBLE REGION

The equation for a normal curve fit to a histogram has a constant factor of $N\,w\,/\,(2\pi)^{(1/2)}$, where N is the number of data points and w is the width of the class intervals used to construct the histogram. For convenience, and without loss of generality, I omit the constant in the following derivations. The bell curve for the population has the equation

$$y_{pop}(z) = \exp(-z^2/2), \tag{A1}$$

which has an area under the curve of $(2\pi)^{(1/2)}$ and a maximum height of one at $z = 0$. The bell curve for a subpopulation of size $\theta = N_{sub}\,/\,N_{pop}$, $0 < \theta \leq 1$, has the equation

$$y_{sub}(z) = (\theta/\sigma)\exp[-(z-\mu)^2/2\sigma^2], \tag{A2}$$

which has an area under the curve of $\theta\,(2\pi)^{(1/2)}$ and a maximum height of θ/σ at $z = \mu$.

For fixed θ and σ, there is a feasible range over which μ can vary without violating the condition that the subpopulation bell curve lie underneath the bell curve of the population. When μ attains a limit of its feasible range for fixed θ and σ, the subpopulation bell curve touches the population bell curve at the contact point. Denote the Z coordinate of the contact point z_c. At the contact point, the heights of the two bell curves are equal, $y_{pop}(z_c) = y_{sub}(z_c)$, or

$$\exp(-z_c^2/2) = (\theta/\sigma)\exp[-(z_c-\mu)^2/2\sigma^2]. \tag{A3}$$

Also, at the contact point, the derivatives (dy/dz) of the two curves are equal:

$$\exp(-z_c^2/2)(-z_c) = (\theta/\sigma)\exp[-(z_c-\mu)^2/2\sigma^2](-1/2\sigma^2)\,2\,(z_c-\mu) \tag{A4}$$

Combining (A3) and (A4) yields

$$-z_c = (-1/2\sigma^2)\,2\,(z_c-\mu). \tag{A5}$$

Multiplying both sides of (A5) by σ^2 and simplifying yields

$$-z_c\,\sigma^2 = -z_c + \mu. \tag{A6}$$

Adding z_c to both sides of (A6), factoring $z_c - z_c\,\sigma^2$ to $z_c(1-\sigma^2)$, and solving for z_c yields

$$z_c = \mu\,/\,(1-\sigma^2), \tag{A7}$$

where, to avoid division by zero, $\sigma = 1$ is not allowed. Substituting $\mu/(1-\sigma^2)$ for z_c in (A3) yields

$$\exp\{-[\mu/(1-\sigma^2)]^2/2\} = (\theta/\sigma)\exp\{-[\mu/(1-\sigma^2)-\mu]^2/2\sigma^2\}. \quad \text{(A8)}$$

Taking natural logarithms of both sides of (A8) gives

$$-[\mu/(1-\sigma^2)]^2/2 = \ln(\theta/\sigma) - [\mu/(1-\sigma^2)-\mu]^2/2\sigma^2. \quad \text{(A9)}$$

Multiplying out the squares within brackets gives

$$-[\mu^2/(1-\sigma^2)^2]/2 = \ln(\theta/\sigma) - [\mu^2/(1-\sigma^2)^2 - 2\mu^2/(1-\sigma^2) + \mu^2]/2\sigma^2 \quad \text{(A10)}$$

Moving the second term of the right side to the left side and factoring out μ^2 gives

$$\mu^2\{-[1/(1-\sigma^2)^2]/2 + [1/(1-\sigma^2)^2 - 2/(1-\sigma^2) + 1]/2\sigma^2\} = \ln(\theta/\sigma) \quad \text{(A11)}$$

The expression within braces, {}, can be simplified. Start by multiplying and dividing by the factors to obtain

$$-1/2(1-\sigma^2)^2 + 1/2\sigma^2(1-\sigma^2)^2 - 2/2\sigma^2(1-\sigma^2) + 1/2\sigma^2. \quad \text{(A12)}$$

Placing all the terms of (A12) on a common denominator gives

$$[-\sigma^2 + 1 - 2(1-\sigma^2) + (1-\sigma^2)^2]/2\sigma^2(1-\sigma^2)^2. \quad \text{(A13)}$$

The first two terms, $-\sigma^2 + 1$, are $1-\sigma^2$, which makes $1-\sigma^2$ a factor common to the numerator and the denominator; canceling out the common factor yields

$$[1 - 2 + (1-\sigma^2)]/2\sigma^2(1-\sigma^2). \quad \text{(A14)}$$

The numerator simplifies to $-\sigma^2$, which cancels the σ^2 in the denominator; so $-1/2(1-\sigma^2)$ is the expression within the braces of (A11). Hence (A11) becomes

$$\mu^2\{-1/2(1-\sigma^2)\} = \ln(\theta/\sigma). \quad \text{(A15)}$$

Multiplying both sides of (A15) by $-2(1-\sigma^2)$ yields

$$\mu^2 = -2(1-\sigma^2)\ln(\theta/\sigma). \quad \text{(A16)}$$

The square root of (A16) is equation (4) of the text.

APPENDIX B

MAXIMUM SUBPOPULATION MEAN

Because the equation

$$\mu = \pm [-2\,(1-\sigma^2)\ln(\theta/\sigma)]^{1/2} \tag{B1}$$

gives the maximum feasible μ as a function of θ and σ, it is possible to obtain, for fixed θ, the σ at which the maximum feasible μ occurs by setting the derivative of (B1) with respect to σ equal to zero, and solving for σ. Monotonic transformations are often used to simplify this process: if μ attains its maximum at $\sigma = \sigma_x$, then any monotonic transformation of μ will have a maximum at $\sigma = \sigma_x$. Thus, squaring (B1) gives

$$\mu^2 = -2\,(1-\sigma^2)\ln(\theta/\sigma). \tag{B2}$$

Dividing by 2, noting that $-1\,(1-\sigma^2) = (\sigma^2 - 1)$, and separating $\ln(\theta/\sigma)$ gives

$$\mu^2/2 = (\sigma^2 - 1)\,[\ln(\theta) - \ln(\sigma)]. \tag{B3}$$

or

$$\mu^2/2 = \sigma^2\ln(\theta) - \sigma^2\ln(\sigma) - \ln(\theta) + \ln(\sigma). \tag{B4}$$

Taking the derivative of (B4) with respect to σ gives

$$d(\mu^2/2)/d\sigma = 2\,\sigma\ln(\theta) - \sigma^2/\sigma - 2\,\sigma\ln(\sigma) + 1/\sigma \tag{B5}$$

Setting (B5) equal to zero fixes the value of σ at σ_x:

$$2\,\sigma_x\ln(\theta) - \sigma_x - 2\,\sigma_x\ln(\sigma_x) + 1/\sigma_x = 0 \tag{B6}$$

Multiplying by $-1/2\,\sigma_x$ gives

$$-\ln(\theta) + 1/2 + \ln(\sigma_x) - 1/(2\,\sigma_x^2) = 0, \tag{B7}$$

which is not readily solved for σ_x as a function of θ. Thus, for fixed θ, I used a binary search (also known as bisection) to bracket σ_x. The initial bounds for σ_x were θ and 1; iteration continued until the difference between the lower bound and the upper bound was less than 10^{-6}. Then (B1) was used to obtain μ_x from σ_x.

Blank

APPENDIX C

EXTREME SUBPOPULATION-TO-POPULATION CONVERSION

When estimating the population ED_{50} from a subpopulation ED_{50} and the subpopulation probit slope, the extreme case for conversion process is not at $\mu = \mu_x$. To find the combination of μ and σ yielding the largest difference between the subpopulation median and the population median, start with the equation for the conversion factor (CF) of a subpopulation ED_{50} to the population ED_{50}:

$$CF = \text{antilog}\,[\,(0 - \mu\,)\,/\,\sigma\, m_{sub}\,]. \tag{C1}$$

Equation (C1) can be rewritten as $CF = \text{antilog}\,[\,(-1\,/\,m_{sub})\,(\,\mu\,/\,\sigma\,)\,]$, from which it is clear that the CF is maximized when the ratio $\mu\,/\,\sigma$ is maximized. To proceed from (C1), take logarithms of both sides of (C1) and substitute $[\,-2\,(1-\sigma^2)\,\ln(\,\theta\,/\,\sigma\,)\,]^{1/2}$ from equation (4) of the text for μ:

$$\log(\,CF\,) = \{\,0 - [\,-2\,(1-\sigma^2)\,\ln(\,\theta\,/\,\sigma\,)\,]^{(1/2)}\}\,/\,\sigma\, m_{sub} \tag{C2}$$

Multiply both sides by m_{sub} and then square both sides:

$$[m_{sub}\log(CF)]^2 = [\,-2\,(1-\sigma^2)\,\ln(\,\theta\,/\sigma\,)\,]\,/\,\sigma^2 \tag{C3}$$

Moving the σ^2 of the denominator into the numerator and rewriting $\theta\,/\sigma$ as $\theta\,\sigma^{-1}$ yields

$$[m_{sub}\ \log(CF)]^2 = \ -2\,(\sigma^{-2} - 1\,)\,\ln(\theta\,\sigma^{-1}) \tag{C4}$$

Now take the derivative with respect to σ and set the derivative equal to zero. Setting the derivative to zero fixes the value of σ, so it is denoted σ_r.

$$-2\,[\,-2(\sigma_r^{-3})\,\ln(\theta\,\sigma_r^{-1}) + (\sigma_r^{-2} - 1\,)\,(-\theta\,\sigma_r^{-2}/\,\theta\,\sigma_r^{-1})\,] = 0 \tag{C5}$$

Simplify $(-\theta\,\sigma_r^{-2}/\,\theta\,\sigma_r^{-1})$ to $(-\sigma_r^{-1})$ and multiply both sides by $-\sigma_r^3/\,2$:

$$-2\,\ln(\theta\,\sigma_r^{-1}) + \sigma_r^3\,(\sigma_r^{-2} - 1\,)\,(-\sigma_r^{-1})\ = 0 \tag{C6}$$

Multiplying out $\sigma_r^3\,(\sigma_r^{-2} - 1\,)\,(-\sigma_r^{-1})$ reduces (C6) to

$$-2\,\ln(\theta\,/\,\sigma_r) + (\sigma_r^{\,2} - 1\,) = 0 \tag{C7}$$

Separating $-2\,\ln(\theta\,/\,\sigma_r)$ to $-2\,\ln(\theta) + 2\,\ln(\sigma_r)$, dividing both sides by 2, and adding $\ln(\theta)$ to both sides produces

$$\ln(\sigma_r) + (\sigma_r^{\,2} - 1)\,/\,2 = \ln(\theta). \tag{C8}$$

Subtracting $(\sigma_r^2 - 1)/2$ from both sides of (C8) followed by exponentiation of both sides yields

$$\sigma_r = \theta \exp[(1 - \sigma_r^2)/2]. \tag{C9}$$

Equation (C9) was solved numerically (by iteration) to obtain σ_r. The starting value for σ_r was $(\theta + 1)/2$. Because an overestimate of σ_r [as the input on the right side of (C9)] yields an underestimate [as the output on the left side of (C9)] and vice versa, the input and output values of σ_r were averaged to obtain the input for the next iteration. Iteration continued until the difference between the input value and the output value was less than 10^{-6}. In Figure 2 of the text, the point (σ_r, μ_r) can be found by drawing a line from the origin tangent to the feasible region; the line will have slope $= \mu_r / \sigma_r$.