

EPA/600/R-00/058

July 2000

<http://www.epa.gov/ncea/>

Options for Development of Parametric Probability Distributions for Exposure Factors

National Center for Environmental Assessment-Washington Office
Office of Research and Development
U.S. Environmental Protection Agency
Washington, DC

DISCLAIMER

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

CONTENTS

	<u>Page</u>
LIST OF TABLES	vi
LIST OF FIGURES	vi
LIST OF ABBREVIATIONS	vii
PREFACE	ix
AUTHORS AND REVIEWERS	x
1. INTRODUCTION	1-1
1.1 Review of Pertinent Statistical Theory and Concepts	1-2
1.2 Maximum Likelihood Estimation	1-7
1.3 Probability Models	1-12
1.4 Assessment of Goodness-of-Fit	1-14
1.5 Uncertainty in Monte Carlo Risk Assessment Models	1-18
1.6 Summary of a System for Fitting Exposure Factor Distributions	1-21
1.6.1 Models	1-21
1.6.2 Methods of Estimation of Model Parameters	1-22
1.6.3 Methods of Assessing Statistical GOF of Probability Models	1-22
1.6.4 Methods of Estimating Uncertainty in the Model Parameters	1-23
1.6.5 System Output	1-23
2. A SYSTEM FOR FITTING DISTRIBUTIONS TO EXPOSURE FACTOR DATA	2-1
2.1 Models	2-1
2.2 Methods of Estimation	2-2
2.2.1 Maximum Likelihood Estimation	2-3
2.2.2 Minimum Chi-Square Estimation	2-3
2.2.3 Weighted Least Squares and Minimum Distance Estimation	2-4
2.2.4 Method of Moments Estimation	2-4
2.2.5 Estimation by Meta-Analysis	2-4
2.2.6 Regression on Age and Other Covariates	2-5
2.2.7 Distributions of Related Test Statistics	2-5
2.2.8 Recommended Methods of Estimation and Discussion	2-6
2.3 Methods of Assessing Statistical Goodness-of-Fit (GOF)	2-6
2.3.1 P-P Plots, Q-Q Plots, and Percent Error Plots	2-7
2.3.2 Relative and Absolute Tests of Model Fit	2-7
2.3.3 Likelihood Ratio Test of Fit Versus a More General Model	2-8
2.3.4 F Test of Fit Versus a More General Model	2-8
2.3.5 Pearson Chi-Square Goodness-of-Fit Test	2-8

CONTENTS (continued)

2.3.6	GOF Tests Based on the EDF	2-9
2.3.7	Recommended Methods for Assessing Statistical GOF and Discussion	2-9
2.4	Methods of Obtaining Distributions for Parameter Uncertainty	2-10
2.4.1	Model Uncertainty	2-10
2.4.2	Parameter Uncertainty	2-11
2.4.2.1	Uncertainty Analysis Based on Asymptotic Normality of Parameter Estimates	2-12
2.4.2.2	Uncertainty Analysis Based on Bootstrapping	2-12
2.4.2.3	Uncertainty Analysis Based on the Normalized Likelihood	2-12
2.4.2.4	Uncertainty Based on Meta-Analysis	2-13
2.4.3	Recommended Method for Uncertainty and Discussion	2-13
2.5	System Output (Summary of Reported Statistics)	2-15
3.	ANALYSIS OF TAP WATER DATA	3-1
3.1	Methods	3-1
3.2	Results	3-2
3.3	Uncertainty Analysis	3-4
3.4	Conclusions	3-5
4.	ANALYSIS OF POPULATION MOBILITY DATA	4-1
4.1	Methods	4-1
4.1.1	Data	4-1
4.1.2	Statistical Methods	4-2
4.2	Results	4-3
4.3	Uncertainty Analysis	4-3
4.4	Conclusions	4-3
5.	APPLICATION TO INHALATION RATES	5-1
5.1	Data	5-2
5.2	Statistical Methods	5-4
5.3	Results	5-5
5.4	Conclusions	5-5
6.	DISCUSSION AND RECOMMENDATIONS	6-1
6.1	Adequacy of Data	6-1
6.2	Application of Methodology to Other Exposure Factors	6-2
6.2.1	Case 1: Percentile Data	6-3
6.2.2	Case 2: Three to Five Statistics Available	6-4
6.2.3	Case 3: Two Statistics Available	6-5

CONTENTS (continued)

6.2.4	Case 4: At Most, One Statistic Is Available	6-6
6.2.5	Topics for Future Research	6-6
7.	REFERENCES	7-1
8.	APPENDICES	
	Appendix A. Glossary and Abbreviations	A-1
	Appendix B. Fitting Models to Percentile Data	B-1
	Appendix C. Fitting Quantiles by Combining Nonlinear and Linear Regression	C-1
	Appendix D. The Generalized (Power Transformed) F Family of Nonnegative Probability Distributions	D-1

LIST OF TABLES

		<u>Page</u>
Table 1-1.	Three MLEs of Prevalence, Given Different Observed Numbers of Infections	1-24
Table 1-2.	Computation of Chi-Square GOF for Tap Water Consumption by 65 Years or Older; Probability Model Is a Gamma Distribution	1-24
Table 3-1a.	Chi-Square GOF Statistics for 12 Age-Specific Models, Fit to Tap Water Data, Based on Maximum Likelihood Method of Parameter Estimation	3-7
Table 3-1b.	P-Values for Chi-Square GOF Tests of 12 Age-Specific Models, Tap Water Data	3-8
Table 3-2.	Results of Statistical Modeling of Tap Water Data Using Five-Parameter Generalized F and Two-Parameter Gamma, Lognormal, and Weibull Models	3-9
Table 3-3.	Uncertainty Distribution of Gamma Parameters Estimated from Tap Water Data.	3-11
Table 3-4.	Results of Two-Step Simulation Process to Incorporate Uncertainty into Drinking Water Distributions Using Asymptotic Normality	3-11
Table 3-5.	Uncertainty Analysis Based on Asymptotic Normality Using Two-Step Simulation Process for Two-Parameter Gamma Distributions	3-12
Table 4-1.	Selected Percentiles of Residence Times in Years from Three Key Studies ...	4-5
Table 4-2.	Residence Time Distributions in Years from Johnson and Capel (1992)	4-6
Table 4-3.	Results of Statistical Modeling of Population Mobility Data.	4-7
Table 5-1.	Parameter Estimates for Individual Factors Affecting Long-Term Inhalation Rates (m^3/day)	5-7
Table 5-2.	Estimated Mean, Coefficient of Variation, and Quantiles for Inhalation Rate (m^3/day), Assuming Gamma or Lognormal Distribution	5-7

LIST OF FIGURES

Figure 1-1.	Histograms and PDFs	1-26
Figure 1-2.	PDFs and CDFs	1-27
Figure 1-3.	Examples of Parametric PDFs	1-28
Figure 1-4.	Demonstration of MLE	1-29
Figure 1-5.	Tap Water Gamma GOF Plots: Adults 65 Years Old and Older	1-30
Figure 1-6.	Tap Water Gamma P-P Plot and Q-Q Plots: Adults 65 Years Old and Older .	1-31
Figure 3-1.	Tap Water Intake P-P Plots, EFH Table 3-7, Children	3-13
Figure 3-2.	Tap Water Intake Q-Q Plots, EFH Table 3-7, Children	3-14

LIST OF ABBREVIATIONS

AD	Anderson Darling
BMR	basal metabolic rate (MJ/day)
CDF	cumulative distribution function
CV	coefficient of variance
CvM	Cramer-von Mises
df	degrees of freedom
EDF	empirical distribution function
EFH	Exposure Factors Handbook
GOF	goodness-of-fit
H	oxygen uptake factor, the volume of oxygen (at standard temperature and pressure, dry air) consumed in the production of 1 MJ energy expended (m ³ /MJ)
iid	identically and independently distributed
KS	Kolmogorov-Smirnov
LRT	likelihood ratio test
MAAPE	minimized average percent error
MCS	minimum chi-square
MDE	minimum distance estimation
ML	maximum likelihood
MLE	maximum likelihood estimator
MOM	method of moments
NFCS	National Food Consumption Survey
NHANES	National Health and Nutrition Examination Survey

LIST OF ABBREVIATIONS (continued)

OLS	ordinary least squares
PAR	population at risk
PDF	probability density function
RA	risk assessment
WLS	weighted least squares
WSE	weighted sum of squares of errors

PREFACE

The National Center for Environmental Assessment (NCEA)–Washington Office within EPA’s Office of Research and Development (ORD) has prepared this document in response to requests from users of the Exposure Factors Handbook (EPA/600/P-95/002Fa-Fc, August 1977) who expressed the need for assistance in using probabilistic methods in exposure assessments. This document summarizes procedures to fit distributions to selected data from the Exposure Factors Handbook.

AUTHORS AND REVIEWERS

The National Center for Environmental Assessment (NCEA)–Washington Office within EPA’s Office of Research and Development was responsible for the preparation of this document. The original document was prepared by the Research Triangle Institute under EPA Contract No. 68D40091, Work Assignment No. 97-12. Jacqueline Moya of NCEA-Washington Office served as the EPA Work Assignment Manager, providing overall direction and coordination of the production effort.

AUTHORS

Lawrence Myers
Michael Riggs
Justin Lashley
Roy Whitmore

Research Triangle Institute
Research Triangle Park, NC

Jacqueline Moya
NCEA–Washington Office
Washington, DC

REVIEWERS

Preliminary drafts of this document received internal and external peer review. Reviewers included:

U.S. Environmental Protection Agency Reviewers:

Hans Allender	Office of Prevention, Pesticides, and Toxic Substances
Tim Barry	Office of Policy
Steven Chang	Office of Emergency and Remedial Response
Henry Kahn	Office of Water
Patrick Kennedy	Office of Prevention, Pesticides, and Toxic Substances

AUTHORS AND REVIEWERS (continued)

Steven Knott	Risk Assessment Forum, National Center for Environmental Assessment
Elizabeth H. Margosches	Office of Prevention, Pesticides, and Toxic Substances
Tom McCurdy	National Exposure Research Laboratory

External Reviewers:

Robert Fares	Environmental Standards Inc.
Paul Price	Ogden Environmental and Energy Services
Curtis Travis	Project Performance Corporation
Alan Stern	New Jersey Department of Environmental Protection
Shree Y. Whitaker	National Institute of Environmental Health Sciences