United States
Environmental Protection
Agency

Office of Monitoring Support and
Quality Assurance
Washington DC 20460

Research and Development

EPA-600/S4-80-026 Oct. 1982

EPA

Project Summary

Correlations Between Age-Adjusted Mortality Rates for White Males and Females in the United States, by County: 1968-1972

Lance Wallace and Valarie J. Gill

This study was undertaken to test the suitability of the User-Prompted Graphics Data Evaluation (UPGRADE) computer system for use in environmental health studies. All possible pairs of correlation (2400 in all) between county mortality rates for about 50 causes of death were systematically investigated, using nearly all of the nation's 3082 counties in each calculation. The causes of death with the highest correlations were determined and their geographic variations mapped.

The strongest correlations were found between diseases that are closely associated with population density. In some cases, diseases affecting related organs were strongly correlated, but in many cases they were not correlated. The common practice of combining death from diseases affecting related organs may thus obscure rather than clarify associations with environmental or other variables.

The distribution of county mortality rates was shown to be neither normal nor log-normal. Thus the interpretation of the absolute values of the correlation coefficients is uncertain; however, their relative rankings may be more trustworthy.

The UPGRADE computer system used to perform the calculations was shown to be useful in providing easy access to a valuable data base, but was also shown to lack certain features that would have made the calculations more efficient and cost-effective.

This Project Summary was developed by EPA's Office of Monitoring Support and Quality Assurance, Washington, D.C., to announce key findings of the research project that is fully documented in a separate report of the same title (see Project Report ordering information at back).

Introduction

A flexible computer graphics and analysis system that can provide simultaneous access to environmental and health data bases would be helpful in environmental health studies. The United States Environmental Protection Agency has partially supported the development of such a system: User-Prompted Graphics Data Evaluation (UPGRADE). As an early test of the system's capabilities, a study was undertaken of correlations between county mortality rates.

A knowledge of correlations between causes of death is useful for several reasons:

1) Strong correlations may imply common factors in the development of the disease. Conversely, weak correlations may be a warning that the diseases are not closely related and that they should not be combined

or otherwise associated in epidemiological studies.

2) Differences or similarities of correlations between sexes or races may indicate genetic or occupational factors of importance in tracing disease etiology.
3) Negative correlations may indicate competing causes of death.

The objective of the study was to evaluate use of the UPGRADE system to calculate all possible correlations, determine the strongest correlations, record them for future use by interested researchers, and investigate the geographical variation of the strongly correlated diseases.

Experimental Procedure.

The data base used in the study consisted of county-level, age-adjusted mortality rates averaged over the five-year period 1968-1972. The rates were calculated by Herb Sauer of the University of Missouri, using the detailed mortality records provided by the National Center for Health Statistics (NCHS). All deaths were recorded between 1968 and 1971, but only every other death in 1972. Thus, some sampling error could exist for the less common diseases. Death rate calculations were based on each county's 1970 population. About 50 causes of death were studied for white males and females (Table 1).

Because the 3082 mortality rates for almost any cause of death contained some 10-30 extraordinarily high rates, due often to confounding factors such as the existence of a major institution (Indian reservation, regional hospital, prison) in the county, and because these rates could exert undue influence on the Pearson correlation coefficient, such outliers were eliminated by use of a scatterplot screening technique. Visual inspection of the scatterplots suggested reasonable upper and lower bounds for county mortality rates to be included in the correlation calculations. Varying these limits provided an indication of the sensitivity of the calculations to the number of counties included: only about a 10-15% variation in the most significant correlations was observed.

A stringent significance criterion of $p < .0001$ was chosen to lessen the likelihood of error in identifying significant correlations. Even so, of the approximately 1200 possible correlations for each sex, 152 correlations were significant for white females and 136 correlations were significant for white males at the $p < .0001$ level.

Table 1. *UPGRADE Variables Used in This Study with Corresponding ICDA Codes*

UPGRADE CODE	*VARIABLE*	*ICDA CODES (8th Revision)*
071	*Tuberculosis, All Forms*	*010-019*
072	*Other Infective Disease*	*000-009, 020-136*
073	*Ca Buccal Cavity, Pharynx*	*140-149*
074	*Cancer of Esophagus*	*150*
075	*Cancer of Stomach*	*151*
076	*Cancer of Intestine*	*152, 153*
077	*Cancer of Rectum*	*154*
078	*Ca Liver, Gall B., Ducts*	*155, 156*
079	*Cancer of Pancreas*	*157*
080	*Other Digestive Cancer*	*158, 159*
081	*Cancer of Resp. System*	*160-163*
082	*Cancer of Breast*	*174*
083	*Cancer of Cervix*	*180*
084	*Cancer of Uterus*	*181, 182*
085	*Ca Prost, Other Female Ca*	*183, 184, 185*
086	*Cancer of Bladder*	*188*
087	*Cancer of Kidney, Etc.*	*189*
088	*Cancer of Central Nervous System (CNS)*	*191, 192*
089	*Residual Cancer*	*170-3, 183, 186-7, 190, 194*
090	*Cancer, Ill-Def. & Sec.*	*195-199*
091	*Lymphosarcoma, Etc.*	*200*
092	*Hodgkin's Disease*	*201*
093	*Multiple Myeloma*	*203*
094	*Leukemia*	*204-207*
095	*Other Lymphatic*	*202, 208, 209*
096	*Neoplasms, Benign & Unspecified*	*210-239*
097	*Diabetes*	*250*
098	*Alcoholism*	*303*
099	*Rheumatic Heart Dis.*	*390-398*
100	*Hypertension*	*400-404*
101	*Acute Ischemic Heart Dis.*	*410, 411*
102	*Chronic Ischemic Heart*	*412, 413*
103	*Other Heart Disease*	*420-429*
104	*Cerebrovascular Disease*	*430-438*
105	*Arteriosclerosis*	*440*
106	*Aortic Aneurysm*	*441*
107	*Other Arteries, Etc.*	*442-448*
108	*Veins, Etc.*	*450-458*
109	*Influenza and Pneumonia*	*470-486*

(continued)

Table 1. *(Continued)*

UPGRADE CODE	VARIABLE	ICDA CODES (8th Revision)
110	Chronic Resp. Dis.	490-493, 517-519
111	Cirrhosis of Liver	571
112	Chronic Nephritis, Etc.	582-584
113	Infections of Kidney	590
114	Congenital Heart & Circ.	746, 747
115	Other Congenital	740-745, 748-759
116	Other Early Infancy	760-778
117	Symptoms, Ill-Defined	780-796
126	Major CV Diseases	390-448
127	Cancer, All Sites and Forms	140-209

Table 2. *Correlation Coefficients for the Top Twenty Correlations*

White Females

Disease Title	Scatterplot Method	Exclusion/ Filter Method
1. Cancer of the Respiratory System - Cirrhosis	.224	.206
2. Cancer of the Intestine - Cancer of the Breast	.211	.161
3. Chronic Ischemic - Cancer, All Forms	.189	.238
4. Chronic Ischemic - Cirrhosis	.182	.198
5. Cancer of the Intestine - Cancer of the Rectum	.180	.145
6. Cancer of the Cervix - Major CV	.180	.189
7. Other Heart Disease - Symptoms, Ill-Defined	.179	.203
8. Rheumatic Heart Disease - Chronic Ischemic	.175	.193
9. Chronic Ischemic - Other Heart	−.172	−.201
10. Acute Ischemic - Cerebrovascular	.171	.173
11. Cancer of the Rectum - Rheumatic Heart	.170	.143
12. Aortic Aneurysm - Cirrhosis	.170	.152
13. Cancer of the Esophagus - Cirrhosis	.169	.118
14. Cancer of the Rectum - Chronic Ischemic	.167	.151
15. Rheumatic Heart - Cirrhosis	.166	.164
16. Diabetes - Major CV	.166	.210
17. Rheumatic Heart - Aortic Aneurysm	.161	.141
18. Cancer of the Rectum - Cancer of the Breast	.159	.155
19. Cirrhosis - Cancer, All Forms	.150	.194
20. Major CV - Cancer, All Forms	.146	.149

The top 30 of these correlations for each sex were further examined. If outliers were suspected, a new modified regression was run using different boundaries for excluding counties. (In most cases, fewer than 1% of all counties were excluded). This procedure resulted in some changes of order among the top correlations, but few sharp changes in the magnitudes of the correlation coefficients.

Results and Discussion

From the procedures discussed above, a final list of the 20 strongest correlations was obtained (Tables 2 and 3). No fewer than eleven correlations appear in both tables, and only two pairs of diseases for each sex were not strongly correlated in the other sex (Table 4). Thus, sex is not a strong factor in the co-variation of mortality rates for most diseases.

However, population density is very clearly an important factor in the most strongly correlated disease pairs, as can be seen by comparing those causes of death most strongly associated with county population to those most strongly correlated with each other. For white females, six of 48 causes of death investigated showed a strong ($p < .0001$) increase in mortality rates in the more populous counties (Table 5). Four of these six appear most often in the strongest 20 correlations for females. Similarly, nine of 46 causes of death investigated for white males showed a strong ($p < .0001$) increase with county population (Table 6). Six of these nine appear most often in the strongest 20 correlations for white males.

The strongest negative correlations are dominated by the "miscellaneous" categories of "Other Heart Disease" and "Symptoms, Ill-defined" (Table 7). These categories probably "compete" with other causes of death in the sense that inexperienced or untrained county medical officers are more likely to classify difficult cases in the miscellaneous category. However, the frequent appearance of rheumatic heart disease in this table does not appear to be explainable in the same way. Rheumatic heart disease appears only for white females and only in association with diseases that have higher mortality rates in rural regions. This phenomenon seems worthy of further study.

The Pearson product-moment correlation coefficient calculation assumes a normal distribution. However, the distribution of county mortality rates was cal-

Table 3. *Correlation Coefficients for the Top Twenty Correlations*

White Males

Disease Title	Scatterplot Method	Exclusion/ Filter Method
1. Other Heart Disease - Symptoms, Ill-Defined	.286	.256
2. Cancer of the Respiratory System - Major Cardiovascular	.286	.302
3. Chronic Ischemic Heart Disease - Aortic Aneurysm	.268	.201
4. Chronic Ischemic - Cirrhosis of the Liver	.263	.221
5. Chronic Ischemic - Cancer, All Forms and Sites	.250	.217
6. Cirrhosis - Aortic Aneurysm	.246	.146
7. Cirrhosis - Cancer, All Forms	.243	.249
8. Cancer of the Respiratory System - Chronic Ischemic	.242	.223
9. Cancer of the Rectum - Cancer of the Intestine	.242	.168
10. Cancer of the Rectum - Chronic Ischemic	.241	.205
11. Major CV - Cancer, All Forms	.239	.220
12. Acute Ischemic Heart Disease - Cerebrovascular	.235	.301
13. Cancer of the Buccal Cavity, Pharynx - Cancer of the Respiratory System	.233	.170
14. Cancer of the Esophagus - Cirrhosis	.231	.186
15. Cancer of the Respiratory System - Chronic Respiratory	.231	.203
16. Cancer of the Respiratory System - Aortic Aneurysm	.226	.173
17. Aortic Aneurysm - Cancer, All Forms	.214	.178
18. Cancer of the Rectum - Cirrhosis	.205	.161
19. Cancer of the Respiratory System - Cancer Ill-Defined and Unspecified	.202	.175
20. Cancer of the Respiratory System - Cirrhosis	.201	.183

Table 4. *Correlations That Are Strong For One Sex But Not The Other*

Rank (WF)		Correlation Coefficient WF	WM
11	Cancer of the Rectum - Rheumatic Heart Disease	.170	.130
16	Diabetes - Major CV Diseases	.166	.098
Rank (WM)			
1	Cancer of the Resp. System - Major CV Diseases	.084	.286
13	Cancer of the Resp. System - Ca. Buccal Cavity	.074	.233

culated for six causes of death for each of three race-sex groups and not one of the 18 data sets passed chi-square tests for normality. In every case, the distributions were more strongly clustered toward the mean and simultaneously more dispersed in the tails than the normal distribution. Such distributions are termed kurtic. The 18 distributions were then plotted on logarithmic probability paper but failed to display log-normal behavior. (Figure 1 provides an example of the nonlinear shape of the distribution). When a more homogenous set of counties is selected, the distribution of mortality rates may more nearly approach log-normality. For example, lung cancer death rates for white males in 234 mostly urban counties were much closer to a log-normal distribution than the rates from all 3082 counties (Figure 2).

Thus we are uncertain of the interpretation to be given to the *absolute* values of the Pearson product-moment correlations calculated in Tables 2 and 3, although the *relative* values may be more trustworthy. For this reason, we have considered only correlations with $p < .0001$. Nonparametric statistics would have been preferable, but because of the large number of counties involved, it was not feasible to calculate Spearman or Kendall correlation coefficients.

It should also be noted that the lack of normality of the county mortality rate distributions probably decreases the allowed range of negative correlations. (For example, two log-normally distributed variables have a minimum r of -0.369, although the positive limit remains at $+1.0$.) Thus, a negative r is probably indicative of a stronger relationship than a positive one of the same magnitude.

Geographic variations were studied using bivariate color maps created by the Domestic Information Display System (DIDS). Rates for each disease were categorized in quartiles, and colors assigned to each of the 16 cells of the resulting 4×4 matrix. Geographic characterizations of six disease pairs showing high correlations for both white males and white females were prepared. An example is given in Table 8.

Two other studies have used similar programs for investigating correlations between diseases. Sauer[1] has grouped the same basic mortality data (1968-72) by state and by state economic area; thus, the present study of county data can be viewed as complementary to Sauer's work. Wellington, MacDonald

Table 5. *Variation in Mortality Rate with County Population (Age-Adjusted Mortality Rate per Million at Risk (1968-72)—White Males)*

Cause of Death	1970 White Male County Population (in thousands) 0-5	5-10	10-25	25-100	>100	P*
Tuberculosis, All Forms	25	30	32	31	35	—
Other Infective Disease	64	65	56	56	53	—
CA Buccal Cavity, Pharynx	45	44	47	54	64	.02
Cancer of Esophagus	26	29	33	38	46	.0001
Cancer of Stomach	94	95	92	92	107	—
Cancer of Intestine	152	157	162	175	208	.0001
Cancer of Rectum	39	46	56	62	72	.0001
Cancer of Liver, Gall B., Ducts	31	28	28	29	35	—
Cancer of Pancreas	110	105	108	108	111	—
Other Digestive Cancer	7	8	8	7	8	—
Cancer of Resp. System	512	529	566	600	645	.0001
Cancer of Breast	3	2	3	3	3	—
Cancer of Prostate	202	195	200	200	198	—
Cancer of Bladder	61	55	59	70	78	—
Cancer of Kidney, Etc.	44	43	43	47	46	—
Cancer of CNS	47	47	49	47	51	—
Residual Cancer	77	77	72	70	68	—
Lymphosarcoma, Etc.	38	36	40	41	43	—
Cancer Ill-Def. and Sec.	109	109	111	113	118	—
Hodgkin's Disease	22	22	21	20	21	—
Multiple Myeloma	24	24	24	25	24	—
Leukemia	108	96	97	92	91	—
Other Lymphatic	25	26	26	25	24	—
Neoplasms, Benign and Unspec.	23	20	22	25	26	—
Diabetes	170	166	171	174	173	—
Alcoholism	24	26	27	25	26	—
Rheumatic Heart Disease	55	61	65	72	83	.0001
Hypertension	117	133	129	121	104	—
Acute Ischemic Heart Dis.	3,006	3,004	3,019	2,860	2,629	.0001
Chronic Ischemic Heart	1,223	1,300	1,445	1,682	1,930	.0001
Other Heart Disease	376	321	293	247	190	.0001
Cerebrovascular Disease	1,207	1,229	1,252	1,166	1,056	.0003
Arteriosclerosis	198	198	202	199	180	—
Aortic Aneurysm	90	92	102	125	129	.0001
Influenza and Pneumonia	415	403	397	387	393	—
Chronic Resp. Disease	403	376	395	423	397	—
Cirrhosis of Liver	115	116	128	158	219	.0001
Chronic Nephritis	44	45	43	47	35	—
Infections of Kidney	50	46	49	44	39	—
Congenital Heart & Cir.	42	44	44	42	41	—
Other Congenital	48	45	47	45	42	—
Other Early Infancy	237	234	228	216	199	.02
Major Cardiovascular Diseases	6,355	6,416	6,587	6,549	6,311	—
Cancer, All Sites and Forms	1,777	1,774	1,848	1,926	2,071	.0001

**Probability that the increase (decrease) in rates is due to chance (Pearson product-moment correlations applied to all counties)*

and Wolf[2] considered cancer mortality between 1950 and 1969 on a statewide basis. Comparisons with results from both works reveals considerable agreement, although different choices of disease groups makes detailed comparisons impossible.

References

1. Sauer, H.I., *Geographic Patterns in the Risk of Dying and Associated Factors: U.S. 1968-72,* National Center for Health Statistics, U.S. Dept. of Health & Welfare, Wash. D.C. 1979.
2. Wellington, MacDonald, and Wolf, *Cancer Mortality: Environmental and Ethnic Factors,* Academic Press, New York, 1979.

Table 6. *Variation in Mortality Rate with County Population (Age-Adjusted Mortality Rate per Million at Risk (1968-72)—White Females)*

Cause of Death	1970 White Female County Population (in thousands)					P*
	0-5	5-10	10-25	25-100	>100	
Tuberculosis, All Forms	9	10	10	10	10	—
Other Infective Disease	56	52	46	42	40	.03
Cancer of Buccal Cavity, Pharynx	16	15	16	17	19	—
Cancer of Esophagus	9	8	9	9	12	.03
Cancer of Stomach	50	46	48	45	52	—
Cancer of Intestine	143	147	156	160	173	.002
Cancer of Rectum	28	29	34	38	40	.002
Cancer of Liver, Gall B., Ducts	37	31	31	32	32	—
Cancer of Pancreas	61	62	64	64	66	—
Other Digestive Cancer	8	6	6	6	6	—
Cancer of Resp. System	85	90	97	107	126	.0001
Cancer of Breast	212	219	228	249	279	.0001
Cancer of Cervix	52	59	61	61	50	—
Cancer of Uterus	42	46	46	44	46	—
Other Female Cancer	81	84	91	95	99	.007
Cancer of Bladder	20	17	19	22	23	—
Cancer of Kidney, Etc.	22	21	21	21	21	—
Cancer of CNS	31	32	33	30	32	—
Residual Cancer	46	44	43	40	39	—
Lymphosarcoma, Etc.	23	24	26	28	28	—
Cancer Ill-Def. and Sec.	85	97	91	89	92	—
Hodgkin's Disease	11	12	11	12	13	—
Multiple Myeloma	17	15	16	17	17	—
Leukemia	58	59	59	56	57	—
Other Lymphatic	16	16	16	15	16	—
Neoplasms, Benign and Unspec.	20	21	18	20	21	—
Diabetes	192	180	187	186	175	—
Alcoholism	6	5	5	6	7	—
Rheumatic Heart Disease	51	50	55	67	82	.0001
Hypertension	104	111	112	97	86	.05
Acute Ischemic Heart Dis.	1,197	1,212	1,202	1,161	1,114	.01
Chronic Ischemic Heart	900	938	1,026	1,179	1,263	.0001
Other Heart Disease	214	195	175	148	137	.003
Cerebrovascular Disease	1,000	978	996	962	899	.009
Arteriosclerosis	162	162	164	168	149	—
Aortic Aneurysm	28	26	28	33	34	—
Influenza and Pneumonia	264	252	243	233	225	.04
Chronic Resp. Disease	89	88	89	95	94	—
Cirrhosis of Liver	51	50	57	74	101	.0001
Chronic Nephritis, Etc.	28	29	29	25	22	—
Infections of Kidney	47	41	42	37	34	—
Congenital Heart & Circ.	33	35	35	34	33	—
Other Congenital	43	41	45	43	40	—
Other Early Infancy	172	161	164	154	143	.04
Major Cardiovascular Diseases	3,712	3,723	3,813	3,869	3,815	—
Cancer, All Sites and Forms	1,153	1,178	1,227	1,254	1,359	.0001

*See note to Table 5.

Table 7. *Strongest Negative Correlations*

			WF	WM
Chronic Ischemic	*vs.*	*Other Heart Disease*	*−.172*	*−.197*
Ca. Rectum	*vs.*	*Other Heart Disease*	*−.105*	*−.146*
Aortic Aneurysm	*vs.*	*Other Heart Disease*	*−.086*	*−.101*
Ca. Breast	*vs.*	*Other Heart Disease*	*−.137*	*NA*
Ca. Intestine	*vs.*	*Other Heart Disease*	*	*−.130*
Acute Ischemic	*vs.*	*Other Heart Disease*		*−.097*
Ca. Intestine	*vs.*	*Other Heart Disease*	*−.125*	
Ca., All Forms	*vs.*	*Other Heart Disease*	*−.123*	
Acute Ischemic	*vs.*	*Symptoms, Ill-Defined*	*−.118*	*−.196*
Chronic Ischemic	*vs.*	*Symptoms, Ill-Defined*	*−.137*	*−.155*
Major CV	*vs.*	*Symptoms, Ill-Defined*	*−.114*	*−.123*
Ca. Rectum	*vs.*	*Symptoms, Ill-Defined*	*−.083*	*−.125*
Ca. All Forms	*vs.*	*Symptoms, Ill-Defined*	*−.139*	
Ca. Breast	*vs.*	*Symptoms, Ill-Defined*	*−.122*	
Ca. Intestine	*vs.*	*Symptoms, Ill-Defined*	*−.123*	
Other Heart Disease	*vs.*	*Rheumatic Heart Disease*	*−.141*	
Infections of Breast	*vs.*	*Rheumatic Heart Disease*	*−.137*	
Symptoms, Ill-Defined	*vs.*	*Rheumatic Heart Disease*	*−.125*	
Acute Ischemic	*vs.*	*Rheumatic Heart Disease*	*−.118*	
Cerebrovascular	*vs.*	*Rheumatic Heart Disease*	*−.092*	
Chronic Isch. Heart Disease	*vs.*	*Cerebrovascular*	*−.141*	*−.093*
Ca. Rectum	*vs.*	*Cerebrovascular*		*−.082*

**Blanks indicate correlations that were not significant at $p \leq .0001$ for the particular sex.*

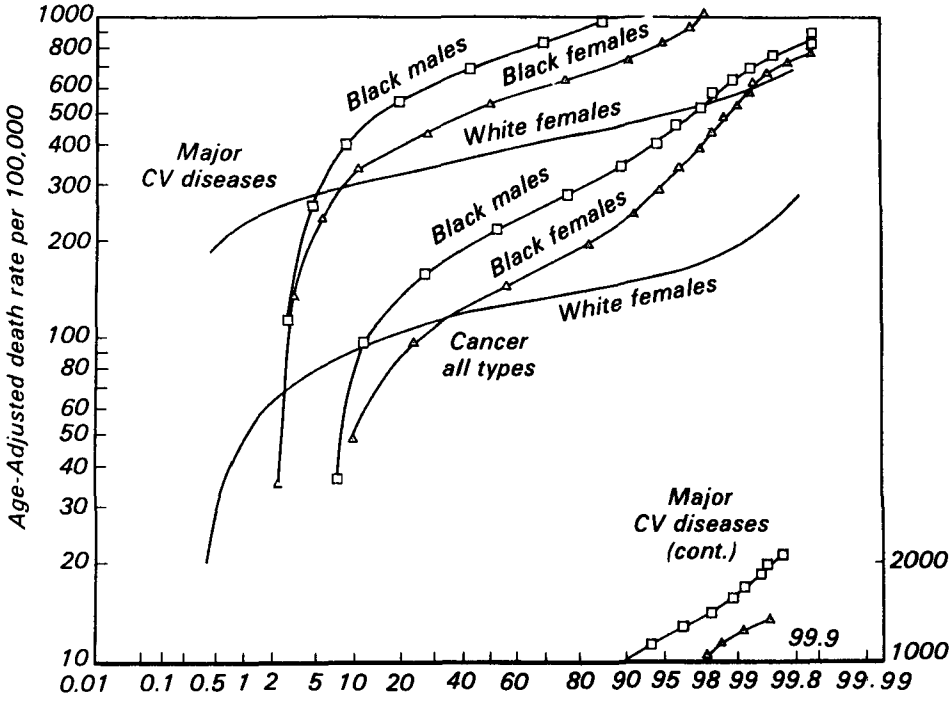

Figure 1. Cumulative frequency distribution of mortality rates: (1968-72).

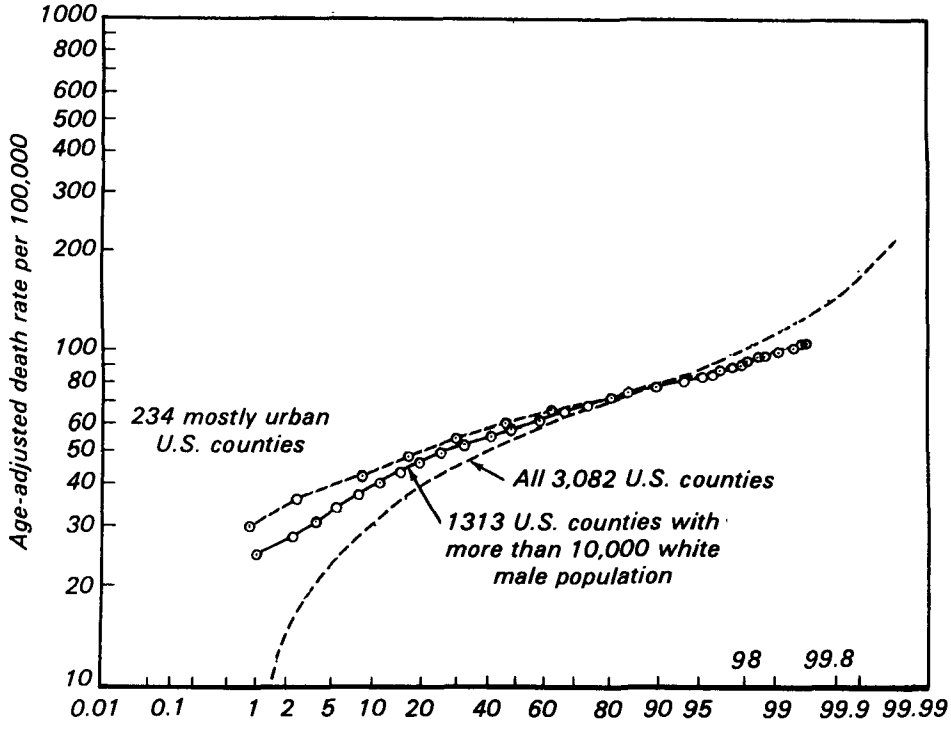

Figure 2. Cumulative frequency distribution of lung cancer mortality rates—white males (1968-72)

Table 8. *Respiratory Cancer vs Cirrhosis of Liver*

Respiratory Cancer	Cirrhosis of Liver	Sex	Geographic Location
HIGH	HIGH	WM	New England, California, Florida
		WF	New England*, California, Florida, Nevada, Arizona, New Mexico, Washington (Seattle-Tacoma-Everett), Gulf Coast, Alaska
LOW	LOW	WM	Tennessee, Kentucky, Virginia
		WF	West**, Southeast
HIGH	MIXED	WM	Georgia, South Carolina, Lower Mississippi River
LOW	HIGH	WM	West, Southwest***

* Particularly CONNECTICUT, MASSACHUSETTS, SOUTHERN VERMONT AND NEW HAMPSHIRE, EASTERN NEW YORK STATE, COASTAL PARTS OF MAIN, MOST OF NEW JERSEY

** Particularly WESTERN PORTION OF NORTH AND SOUTH DAKOTA, NEBRASKA, KANSAS, SOUTHERN PORTION OF MONTANA, IDAHO, UTAH

*** Particularly NEW MEXICO, COLORADO, WYOMING

*The EPA authors **Lance Wallace** and **Valerie J. Gill** (also the EPA Project Officer, see below) are with the Office of Monitoring Support and Quality Assurance, Washington, DC 20460.*

The complete report, entitled "Correlations Between Age-Adjusted Mortality Rates for White Males and Females in the United States, by County: 1968-1972," (Order No. PB 82-224 114; Cost: $10.50, subject to change) will be available only from:

National Technical Information Service
5285 Port Royal Road
Springfield, VA 22161
Telephone: 703-487-4650

The EPA Project Officer can be contacted at:

Office of Monitoring Support and Quality Assurance
U.S. Environmental Protection Agency (RD-680)
Washington, DC 20460

*USGPO:1982-659-095-543

Agency

Cincinnati OH 45268

Environmental
Protection
Agency
EPA-335

Official Business
Penalty for Private Use, $300