NATIONAL BUREAU OF STANDARDS REPORT

10 654

A MODEL TO ESTIMATE THE INCIDENCE OF LEAD PAINT POISONING

NBS

U.S. DEPARTMENT OF COMMERCE

NATIONAL BUREAU OF STANDARDS

NATIONAL BUREAU OF STANDARDS

The National Bureau of Standards[1] was established by an act of Congress March 3, 1901. The Bureau's overall goal is to strengthen and advance the Nation's science and technology and facilitate their effective application for public benefit. To this end, the Bureau conducts research and provides: (1) a basis for the Nation's physical measurement system, (2) scientific and technological services for industry and government, (3) a technical basis for equity in trade, and (4) technical services to promote public safety. The Bureau consists of the Institute for Basic Standards, the Institute for Materials Research, the Institute for Applied Technology, the Center for Computer Sciences and Technology, and the Office for Information Programs.

THE INSTITUTE FOR BASIC STANDARDS provides the central basis within the United States of a complete and consistent system of physical measurement; coordinates that system with measurement systems of other nations; and furnishes essential services leading to accurate and uniform physical measurements throughout the Nation's scientific community, industry, and commerce. The Institute consists of a Center for Radiation Research, an Office of Measurement Services and the following divisions:

> Applied Mathematics—Electricity—Heat—Mechanics—Optical Physics—Linac Radiation[2]—Nuclear Radiation[2]—Applied Radiation[2]—Quantum Electronics[3]—Electromagnetics[3]—Time and Frequency[3]—Laboratory Astrophysics[3]—Cryogenics[3].

THE INSTITUTE FOR MATERIALS RESEARCH conducts materials research leading to improved methods of measurement, standards, and data on the properties of well-characterized materials needed by industry, commerce, educational institutions, and Government; provides advisory and research services to other Government agencies; and develops, produces, and distributes standard reference materials. The Institute consists of the Office of Standard Reference Materials and the following divisions:

> Analytical Chemistry—Polymers—Metallurgy—Inorganic Materials—Reactor Radiation—Physical Chemistry.

THE INSTITUTE FOR APPLIED TECHNOLOGY provides technical services to promote the use of available technology and to facilitate technological innovation in industry and Government; cooperates with public and private organizations leading to the development of technological standards (including mandatory safety standards), codes and methods of test; and provides technical advice and services to Government agencies upon request. The Institute also monitors NBS engineering standards activities and provides liaison between NBS and national and international engineering standards bodies. The Institute consists of the following divisions and offices:

> Engineering Standards Services—Weights and Measures—Invention and Innovation—Product Evaluation Technology—Building Research—Electronic Technology—Technical Analysis—Measurement Engineering—Office of Fire Programs.

THE CENTER FOR COMPUTER SCIENCES AND TECHNOLOGY conducts research and provides technical services designed to aid Government agencies in improving cost effectiveness in the conduct of their programs through the selection, acquisition, and effective utilization of automatic data processing equipment; and serves as the principal focus within the executive branch for the development of Federal standards for automatic data processing equipment, techniques, and computer languages. The Center consists of the following offices and divisions:

> Information Processing Standards—Computer Information—Computer Services—Systems Development—Information Processing Technology.

THE OFFICE FOR INFORMATION PROGRAMS promotes optimum dissemination and accessibility of scientific information generated within NBS and other agencies of the Federal Government; promotes the development of the National Standard Reference Data System and a system of information analysis centers dealing with the broader aspects of the National Measurement System; provides appropriate services to ensure that the NBS staff has optimum accessibility to the scientific information of the world, and directs the public information activities of the Bureau. The Office consists of the following organizational units:

> Office of Standard Reference Data—Office of Technical Information and Publications—Library—Office of International Relations.

---

[1] Headquarters and Laboratories at Gaithersburg, Maryland, unless otherwise noted; mailing address Washington, D.C. 20234.
[2] Part of the Center for Radiation Research.
[3] Located at Boulder, Colorado 80302.

NATIONAL BUREAU OF STANDARDS REPORT

NBS PROJECT		NBS REPORT
4314518	December 7, 1971	10 654

A MODEL TO ESTIMATE THE INCIDENCE OF LEAD PAINT POISONING

Milestone 4

by
Judith Gilsinn
Lambert Joel
Applied Mathematics Division
Michael Thomas
Technical Analysis Division
Lead Paint Poisoning Project
Building Research Division
Institute for Applied Technology
National Bureau of Standards
Washington, D.C. 20234

Sponsored by

Department of Housing and Urban Development

U.S. DEPARTMENT OF COMMERCE

NATIONAL BUREAU OF STANDARDS

PREFACE

This report can be grouped, conceptually, with 4 other NBS Reports under the blanket title "Pediatric Lead Paint Poisoning in the United States--A Survey with Preliminary Estimates." Under this arrangement, the "parts" of the study would be listed:

Part I NBS Report # 10499 "The Nature of the Lead Paint Poisoning Hazard"

Part II NBS Report # 10657 "Data Collection and Assimilation for the Lead Paint Poisoning Model"

Part III NBS Report # 10653 "Effect of Data Aggregation in Modelling"

Part IV NBS Report # 10654 "A Model to Estimate the Incidence of Lead Paint Poisoning"

Part V NBS Report # 10651 "National Estimates of Lead Based Paint Poisoning of Children (Estimated by Standard Metropolitan Statistical Area)"

These papers were intended as interim progress reports covering work done up to the time of publication. Reports describing validation and refinement of the models and outputs as well as analysis of the outputs will be issued subsequently. A summary report encompassing revision of the current ones and those projected above, under one cover, is anticipated.

The bibliography for "Part I" NBS Report # 10499 is appended herein. Numerical references in the body of the report apply to this bibliography. References not appearing in this listing will be found as footnotes marked *. An informal glossary of census related terminology appears at the beginning of the text.

ABSTRACT

This report describes the development of a mathematical model to estimate the incidence in the United States, of pediatric lead poisoning resulting from the ingestion of lead based paint. The kind of model, its purpose, relevant variables and the model form are discussed. Several possible models are listed and compared based on goodness of fit to known data and on the "reasonableness" of their parameter values. The report includes a glossary and a bibliography.

Glossary of Census Related Terminology

The definitions are not all taken verbatim from census sources but are intended to convey the substance of the technical definitions.

Census first count, second count, etc. Census data are assembled and issued in stages. The "first count" is generally information from a 100% sampling and is available for distribution shortly after the end of the nominal year of the decennial census. Fourth count tabulations are developed from smaller samples and might not be published until up to three years after the census data.

Census Tract. A roughly rectangular area with a population of about 5000. Urban census tracts are composed of "block groups", while finer rural subdivisions are called "enumeration districts".

SMSA; Standard Metropolitan Statistical Area. Essentially, a city of at least 50,000 people with its surrounding suburbs. There were 189 tabulated in the 1960 Decennial Census, 243 in 1970.

Household. A group, primarily a family, "living under the same roof".

Crowded Population. Households occupying housing units in which the average number of people per room equals or exceeds fixed thresholds. The two categories which have been tabulated in the census use the threshold values 1.01 and 1.51. This definition does not apply to people in "group quarters"--dormitories, rooming houses, etc.

Housing Categories:

Sound - Having no defects that could not be alleviated by routine maintenance procedures.

Deteriorating - Containing some serious defects: holes in wallboards several inches in diameter, some rotting in structural members, heavy splintering, etc.

Dilapidated - Severe surface damage, fractures in structural members, etc.

Unsound - Deteriorating or Dilapidated.

Substandard - Unsound or lacking an essential plumbing facility.

Essential Plumbing Facility - Hot running water, flush toilet, bath or shower, for the exclusive use of a household.

Dwelling Unit; Housing Unit - Living quarters for a household; primarily a house or apartment. For our purposes the terms are interchangeable. Census used "Dwelling Unit" through 1960 and "Housing Unit" with minor technical changes in the definition, thereafter.

Housing Stock - Housing existing at a given time. The term is used when needed to avoid ambiguity: "Pre 1940 Housing" could mean all housing built before 1940, while "pre 1940 Housing Stock" would be limited to housing built before 1940 and presently standing.

TABLE OF CONTENTS

A MODEL TO PREDICT THE INCIDENCE OF LEAD PAINT POISONING

1. INTRODUCTION

It has been estimated that pediatric lead poisoning resulting from the ingestion of lead based paint is a problem of epidemic proportions in the United States.[27,28] Several cities, notably New York[25] and Chicago* have mass screening programs to identify those children with the disease, but many areas have no screening programs and authorities in some are still unaware of the problem.† Some cities do not even require reporting of lead poisoning cases,[27,28] while others report several hundred cases a year.[25] At present, the number of cases reported appears to depend primarily on the effort spent looking for lead poisoning, rather than on the size of the city. Thus it is desirable that the Federal Government estimate the magnitude and extent of lead poisoning in all areas of the country, in order to determine what level of Federal assistance is most appropriate to combat this problem. This report describes the factors considered by members of the Technical Analysis and Applied Mathematics Divisions of the National Bureau of Standards in an attempt to construct mathematical models with which to estimate the number of children afflicted with lead poisoning from paint ingestion, in cities or roughly equivalent geographical subdivisions and in the nation as a whole.

---

*Conversations with Dr. Herbert Slutsky, Chicago Board of Health.

†Chicago Daily News, 8/3/71.

We are faced with an attempt to develop a forecast model under a limited allocation of time and other resources. We have, moreover, a substantially frozen data base, i.e., there is no possibility of undertaking a structured major program of data acquisition. Such circumstances uncover four question areas which call for early decisions, seriatim, as each decision will influence the nature of the subsequent questions.

The first involves specifying as explicitly as possibly the purpose and scope of the model, because they exert a strong influence on the form of the model and the difficulty of successful completion.

The second general question is: What "kind" of model can be produced?, i.e., what is the major component of the methodological basis for the model? The answers here will determine whether the bulk of the analysis will be qualitative, involving development of theories, or quantitative, relying mainly on manipulation of numerical data.

Third, to what extent can the relevant variables be identified at the outset?

Fourth, what is the most appropriate detailed formal structure of the model in the context of answers to the previous questions?

We note in passing that when a model is to be produced by a group of analysts in response to a request by another person or group, the answer to the first question is primarily the responsibility of this "customer" for the model effort.

2. THE PURPOSE OF THE MODEL

There are two general purposes for developing a model. In one case some output of information is desired and the model is a device for obtaining this output. In the second case the model itself is the end, either because it is to be used to increase the comprehension of a process or system by experimental simulation or because the model is part of a generalized scheme and will be turned over to others who will apply it to specific problems. The model described here falls into the first category. The primary purpose for the lead paint poisoning model is to estimate the present incidence of lead poisoning in the United States. The emphasis is on the model output, rather than the model itself.

This decision on the purpose of the model has made the task easier in several ways, and thus facilitated the construction of a lead poisoning model within the required time span. The model is to be applied by the analysts who designed it. This means that a different kind of documentation is required, one aimed at a proper technical record of what was done, rather than, e.g., a users manual to describe to city officials how to apply the model in their cities.

The predictor variables are the ones which generally are most associated with lead poisoning, although in a specific case other variables might have stronger effect. (As an example, many transportation demand models predict the number of trips between two geographical zones as a function of the trip time, trip cost, and the populations of the two zones. However, in any particular instance the factor which is most important in the trip decision may be none of these, e.g., baggage carrying facilities.) Local conditions vary, but because the model is to be applied at a fairly gross level many of these local variations are subsumed in averages over the larger areas. The fraction of dilapidated or deteriorating housing for an SMSA is likely to have changed less than in a particular census tract. Urban renewal can completely change the character of the housing stock in a particular census tract, but even the largest renewal efforts affect a relatively small proportion of the total number of housing units in a metropolitan area. Thus there are advantages in predicting at this gross level. This is particularly true since some desirable tract data items are available only from the 1960 census and are not yet or will never be available for 1970.

Another advantage of designing a model for producing a particular output is that the application and the characteristics of that application are known in advance. One knows the level of aggregation involved, and thus the ranges of the numbers which will be used. This means that one does not need to be concerned with all possible values of the variables, but only those in the known ranges. The problems associated with different levels of aggregation are reduced since the desired level of aggregation is known at the outset, and problems arising during the model development process can be anticipated and tackled as specific problems when they occur. On the other hand, once having delimited the range of application we cease expending resources on wringing the last dram of flexibility from a formulation.

The purpose of the model described here is to predict the incidence of lead poisoning. The model, which we hope will be valid for cities or SMSA's and possibly census tracts, is not designed to be applied at the micro level to predict for very small areas within a city. It will certainly not be able to predict individual cases. The model as a tool, a part of the prediction process, is intended to furnish us specific information, and has not been designed to facilitate its application in settings merely analogous to its explicit purpose. Further study and testing may indicate that this can be done, but such testing is not part of the current efforts.

3. THE KIND OF MODEL

There are basically two ways to formulate mathematical forecast models. The first of these could be called "cause and effect" modelling. This procedure involves a working hypothesis and can be followed only when the phenomenon being predicted is well enough understood that the modeller can formulate mathematically the "laws" which govern the phenomenon's production as a consequence of its causes. Such a model can then be applied in widely differing contexts by modifying the process equations according to knowledge of the various milieus. However, good working hypotheses for answering socio-economic or demographic questions are hard to come by, since the phenomena under study are seldom well enough understood to make this type of model possible. The second formulation procedure is based essentially on calculations of correlations among sets of historical data or experimental observations. For this reason we will call the result a "correlation" model. This type is used for phenomena for which one knows or suspects several causative factors, but one does not know the precise interrelationships among them and between this causative set and the phenomenon to be predicted. In this case, the exact functional form of the model is not chosen because of a known relationship, but rather in general, to reflect some desired model behavior.

The set of possible model variables may be large and is reduced by comparing model results and fits. Care must be taken in applying such models since the fact that a model fits a particular set of data points does not insure that it will fit others, particularly those out of the range of the data base of observations. Care must also be taken in interpreting such models since the function does not necessarily represent a cause and its effect, but represents a relationship deduced from collections of numbers, a dependence between "dependent" and "independent" variables.

The model described in this report is of the second type. Because of the nature of our current knowledge of lead poisoning, we reiterate that the model described here is a "correlation" model means that care must be taken in applying it, particularly in a context slightly different from the present one. The direct cause of most pediatric lead poisoning is of course the ingestion of lead based paint chips by young children. Authorities disagree on why children eat paint. One group believes that pica, an unreasonable craving for non-food substances, is a major factor.[37] However, the causes of pica are insufficiently understood to be modelled. In fact, even the factors associated with pica are not agreed upon, and many (such as emotional disturbance) are not directly measured by demographic data and would have to be modelled themselves. Thus with little knowledge of the actual causes of paint eating by children, it is impossible to build a "cause and effect" model. However, many factors <u>associated</u> with a high risk of affliction by lead poisoning are known. These include extreme youth (age 6 and under), poverty, low educational level of the mother, female household head, crowding, and living in rented older housing in poor condition. These factors are highly correlated and interrelated. Other factors correlated with the ones listed above may be used to fill in current data gaps when necessary.

4. IDENTIFICATION OF RELEVANT VARIABLES

The process of identifying the variables most relevant for use in the model should, stagewise, choose the dependent variable, identify associated independent variables, and analyze the independent variables.

4.1. The Dependent Variable

The choice of the dependent variable is critical to the modelling process, since it determines what is being modelled. However, this does not mean that the choice is obvious. There were several possible dependent variables for the lead poisoning model:

a. number of lead poisoning cases

b. number of children with elevated blood lead levels (EBL)

c. the percent or fraction of the number of children with elevated blood lead levels

d. the number or proportion of housing units offering a hazard.

Although this last is necessary for estimating costs as a part of planning a deleading program, the primary measure of the lead poisoning problem is the number of children who have lead poisoning. The definition of a lead poisoning "case" is, however, sufficiently vague and differs from place to place and doctor to doctor, so that quantitatively it is meaningless. The output desired from the lead poisoning model is the number of children with lead poisoning, but this can either be predicted directly or by first predicting the incidence rate and applying this rate to the total child population. Both of these methods have been used in the modelling effort for this project, and the one used in a particular model will be indicated in the description of the model. The actual numbers predicted, whether they represent numbers of children or the fraction of children, refer to those with "elevated blood lead levels." The cutoff used here is 40 μg lead per 100 ml. of whole blood, the Surgeon General's recommended level indicating undue absorption of lead.[18] Not all children with blood lead levels above 40 μg/100 ml. require treatment, but this can be determined only by further diagnosis by a doctor. This cutoff level thus distinguishes children who require scrutiny in addition to measurement of blood lead levels even in the absence of accompanying symptomology. Use of a higher cutoff might miss some lead poisoning cases and, what is more important, as is indicated above, would suppress information needed for estimation of total diagnostic costs. The distribution of blood lead levels above 40 μg/100 ml. can be estimated from existing blood lead data so that fractions of the child population with higher levels are also potentially available.

4.2. The Independent Variables

Table 1 contains a list of possible independent variables, all of which have been tabulated at one time or another by the Bureau of the Census. Some of the Census data also contain composites of these variables, such as "female household heads living in crowded housing", or "population under age 18 living in crowded housing". Such composites are especially useful for focusing attention on the particular population of interest,

Table 1: Variables for Use in a Lead Poisoning Model

1. Housing stock built before 1940, and housing stock built between 1940 and 1949.
2. Dilapidated H.U. (housing units)
3. Deteriorating H.U.
4. Value of owner occupied housing (Median dollar value of housing unit e.g., house, coöp apartment, etc.)
5. Monthly contract rent ($/mo.)
6. Housing vacancy rate
7. Multiple unit structures
8. Children under 5 years old
9. Children 5 years old
10. Children 6 years old
11. Female household heads
12. Crowded population ($\geq$ 1.01 per room)
13. Crowded population ($\geq$ 1.51 per room)
14. Region of birth of household head (distribution)
15. Educational level of household head (distribution by years of school completed)
16. Median family income ($/ann.)
17. Number of families with income below the poverty level (Criterion is frequently revised, and is different for rural and urban population)
18. Work status of head of household (Employed or unemployed; distributions by type of employment)
19. Race

since they enumerate people with combinations of the desired characteristics. The sources of these data and the motivation for their use are discussed more fully in NBS Reports #10499, Nature of the Lead Paint Poisoning Hazard, and #10657, Data Collection and Assimilation for the Lead Paint Poisoning Model. A typical lead poisoned child is less than six years old, comes from an urban poor family, and lives in a housing unit containing surfaces with peeling lead based paint. The variables listed in Table 1 help to characterize the areas in which this situation is most likely to exist.

4.2.1. Rankings of Independent Variables

Once the data have been collected and assimilated, some preliminary analysis is carried out. Table 2 lists for 189 SMSA's, the values of several data items and the ranks of the SMSA's by these items. For example, in Abilene, Texas .3998 of the houses were built before 1940; Abilene ranks 41st among the SMSA's in this attribute. All the ranks are from lowest to highest, so that the SMSA ranked 1 in a column has the smallest fraction of houses built before 1940, the smallest fraction of housing rented, the smallest population, or the least average rent. The purpose of this exercise is to display the data in a form to enable one to see plainly trends and relationships between variables. For instance, it can be deduced from Table 2 that pre 1940 housing is concentrated in the Northeast and Mississippi Valley, so that this variable may explain some of the regional differences* which have been associated with lead poisoning incidence.

---

*Conversations with Jim Simpson of the Bureau of Community and Environmental Management, Department of Health, Education, and Welfare.

COLUMN HEADINGS FOR TABLE 2

Note: The 189 SMSA's listed here are those that were SMSA's in 1960.

1. Alphabetical Index of SMSA According to 1970 Listing
2. Name of SMSA
3. "% Old Houses" - Fraction of 1967 Housing Stock Built Before 1940
4. "% Rental" - Fraction of 1970 Housing Stock Renter Occupied or "Vacant for Rent"
5. "Pop" - 1970 Population
6. "Rent" - 1970 Monthly Rent in Dollars
7. "Age of House" - Rank with Respect to Col. 3, (Low Rank ∿ Fewer Pre 1940 H.U.)
8. "% Rented" - Rank with Respect to Col. 4, (Low Rank ∿ Fewer Renters, More Owners)
9. "Pop" - Rank with Respect to Col. 5, (Low Rank ∿ Low Pop)
10. "Rent" - Rank with Respect to Col. 6, (Low Rank ∿ Low Rent)

TABLE 2

		% OLD HOUSES	% RENTAL	POP.	RENT	AGE OF HOUSE	% RENTED	POP.	RENT
1	ABILENE, TEXAS	.3998	.3336	113959.	61.	41	79	1	9
2	AKRON, OHIO	.5923	.2346	679239.	97.	116	29	143	145
4	ALBANY-SCHENECTADY-TROY, NEW YORK	.8701	.3636	721910.	81.	186	117	146	80
5	ALBUQUERQUE, N. MEXICO	.2225	.3472	315774.	86.	5	100	96	102
6	ALLENTOWN-BETHLEHEM-EASTON, PA.-N.J.	.7076	.2928	543551.	79.	166	35	133	66
7	ALTOONA, PENNSYLVANIA	.8231	.2647	135356.	55.	185	13	15	5
8	AMARILLO, TEXAS	.3349	.3202	144396.	76.	21	65	20	62
11	ANN ARBOR, MICHIGAN	.5249	.4285	234103.	142.	79	165	64	187
13	ASHEVILLE, NORTH CAROLINA	.5819	.2872	145056.	57.	109	31	21	12
14	ATLANTA, GEORGIA	.4034	.4250	1390164.	104.	42	162	171	157
15	ATLANTIC CITY, NEW JERSEY	.6747	.3787	175043.	90.	158	128	41	123
16	AUGUSTA, GEORGIA	.4291	.3540	253460.	64.	52	108	69	29
17	AUSTIN, TEXAS	.3551	.4508	295516.	104.	28	172	89	158
18	BAKERSFIELD, CALIFORNIA	.3292	.4048	329162.	72.	20	150	101	53
19	BALTIMORE, MARYLAND	.5510	.4180	2070670.	96.	95	159	179	140
20	BATON ROUGE, LA.	.3061	.3371	285167.	83.	16	88	82	91
21	BAY CITY, MICHIGAN	.6239	.1872	117339.	82.	132	1	5	86
22	BEAUMONT-PORT ARTHUR-ORANGE, TEXAS	.4203	.3038	315943.	67.	48	44	97	36
25	BINGHAMPTON, NEW YORK	.7057	.3074	302672.	84.	165	51	92	95
26	BIRMINGHAM, ALABAMA	.5281	.3366	739274.	57.	83	84	147	13
29	BOSTON, MASSACHUSETTS	.7779	.4742	2753700.	115.	178	181	182	169
30	BRIDGEPORT, CONN.	.5878	.3706	389153.	104.	112	120	115	159
32	BROCKTON, MASSACHUSETTS	.7318	.3151	189820.	88.	173	55	47	115
33	BROWNSVILLE-HARLINGEN-SAN BENITO, TEXAS	.3975	.3261	140368.	48.	40	73	17	2
35	BUFFALO, NEW YORK	.6700	.3706	1349211.	81.	155	121	167	81
36	CANTON, OHIO	.6135	.2652	372210.	78.	127	14	112	67
37	CEDAR RAPIDS, IOWA	.6298	.2813	163213.	96.	136	25	33	141
38	CHAMPAIGN-URBANA, ILLINOIS	.5445	.4610	1632[illegible]1.	94.	92	175	34	142
39	CHARLESTON, SOUTH CAROLINA	.4269	.3987	303849.	69.	51	144	93	41
40	CHARLESTON, WEST VIRGINIA	.5351	.3559	229515.	63.	60	111	61	26
41	CHARLOTTE, NORTH CAROLINA	.3900	.3942	409370.	81.	32	134	116	82
42	CHATTANOOGA, TENN.	.5054	.3379	304927.	64.	76	89	95	30
43	CHICAGO, ILLINOIS	.6891	.4710	6978947.	119.	163	179	187	176
44	CINCINNATI, OHIO-KY.-IND.	.6698	.3902	1384851.	84.	153	139	170	96
45	CLEVELAND, OHIO	.6352	.3760	2064194.	103.	138	125	178	154
46	COLORADO SPRINGS, COLORADO	.4264	.4121	235972.	103.	50	156	66	155
48	COLUMBIA, SOUTH CAROLINA	.3901	.3266	322880.	68.	39	74	99	38
49	COLUMBUS, GEORGIA	.4037	.4727	238544.	62.	43	180	67	23
50	COLUMBUS, OHIO	.5345	.4094	916228.	93.	86	152	156	133
51	CORPUS CHRISTI, TEXAS	.2812	.3554	284832.	70.	13	110	81	47
52	DALLAS, TEXAS	.3545	.3799	1555950.	111.	27	147	174	166
54	DAVENPORT-ROCK ISLAND-MOLINE, IOWA	.6453	.3021	362638.	87.	141	39	111	108
55	DAYTON, OHIO	.5024	.3356	850266.	92.	74	83	152	129
56	DECATUR, ILLINOIS	.6188	.2843	125010.	79.	130	28	10	68
57	DENVER, COLORADO	.4451	.3853	1227529.	107.	59	135	164	162
58	DES MOINES, IOWA	.6249	.3041	286101.	101.	133	47	83	151
59	DETROIT, MICHIGAN	.5275	.2790	4199931.	104.	81	23	185	160
61	DULUTH-SUPERIOR, MINN.-WIS.	.7173	.2673	265350.	66.	169	16	76	34
62	DURHAM, NORTH CAROLINA	.5177	.4606	190388.	79.	78	174	48	71
63	EL PASO, TEXAS	.3857	.4125	359291.	71.	35	157	110	50
64	ERIE, PENNSYLVANIA	.6596	.2843	263654.	71.	149	27	74	51
65	EUGENE, OREGON	.3528	.3592	213358.	89.	26	114	56	120
66	EVANSVILLE, IND.-KY.	.5951	.3023	232775.	70.	118	41	63	48
67	FALL RIVER, MASS.-R.I.	.8049	.5168	149976.	61.	187	185	25	20
68	FARGO-MOREHEAD, N. DAK.-MINN.	.6011	.3752	120238.	92.	123	123	6	130
71	FLINT, MICHIGAN	.5264	.2225	496658.	100.	80	4	124	148
72	FORT LAUDERDALE-HOLLYWOOD, FLORIDA	.0895	.2726	620100.	133.	1	21	137	184
74	FORT WAYNE, INDIANA	.5957	.2695	280455.	94.	119	19	80	134

75	FORT WORTH, TEXAS	.3445	.3337	762086.	97.	24	80	148	146
76	FRESNO, CALIFORNIA	.4082	.3991	413053.	75.	44	145	121	60
79	GALVESTON-TEXAS CITY, TEXAS	.4317	.3753	169812.	74.	55	124	37	57
80	GARY-HAMMOND-EAST CHICAGO, INDIANA	.5327	.3184	633367.	87.	84	62	139	109
81	GRAND RAPIDS, MICHIGAN	.6254	.2274	539225.	83.	134	7	130	116
83	GREEN BAY, WISCONSIN	.5897	.2676	158244.	86.	114	17	31	103
84	GREENSBORO-WINSTON-SALEM-HIGH POINT,N.	.4514	.3383	603895.	69.	61	91	135	42
85	GREENVILLE, SOUTH CAROLINA	.4375	.3192	299502.	57.	57	63	91	14
86	HAMILTON-MIDDLETOWN, OHIO	.5578	.3069	226207.	84.	99	50	58	97
87	HARRISBURG, PENNSYLVANIA	.6430	.3171	410626.	76.	140	59	119	63
88	HARTFORD, CONNECTICUT	.5719	.4096	663891.	117.	103	153	142	173
89	HONOLULU, HAWAII	.3665	.5498	629176.	124.	29	187	138	180
90	HOUSTON, TEXAS	.3062	.3997	1985031.	100.	17	146	177	149
91	HUNTINGTON-ASHLAND, W. VA.-KY.-OHIO	.6348	.3205	253743.	56.	137	66	70	7
92	HUNTSVILLE, ALABAMA	.3371	.3211	228239.	67.	22	67	59	37
93	INDIANAPOLIS, INDIANA	.6071	.3459	1109882.	95.	126	98	162	137
94	JACKSON, MICHIGAN	.6676	.2106	143274.	87.	152	2	19	110
95	JACKSON, MISSISSIPPI	.3666	.3367	258906.	64.	30	85	72	31
96	JACKSONVILLE, FLORIDA	.3791	.3240	528865.	81.	31	71	127	83
97	JERSEY CITY, NEW JERSEY	.9163	.7021	609266.	100.	189	189	136	150
98	JOHNSTOWN, PENNSYLVANIA	.7541	.2888	262822.	47.	175	32	73	1
99	KALAMAZOO, MICHIGAN	.5732	.2952	201550.	108.	104	30	50	163
100	KANSAS CITY, MO.-KANSAS	.5854	.3430	1253916.	95.	111	97	165	138
101	KENOSHA, WISCONSIN	.5882	.2991	117917.	88.	113	38	6	117
102	KNOXVILLE, TENNESSEE	.4638	.3174	400337.	63.	63	61	117	27
106	LAKE CHARLES, LA.	.2757	.2901	145415.	60.	11	33	23	17
107	LANCASTER, PENNSYLVANIA	.6958	.3114	319693.	69.	164	53	98	43
108	LANSING, MICHIGAN	.6057	.3021	378423.	117.	124	40	114	174
110	LAS VEGAS, NEVADA	.1169	.4205	273288.	133.	2	160	78	185
111	LAWRENCE-HAVERHILL, MASS.-N.H.	.7970	.4341	232415.	82.	181	167	62	87
114	LEXINGTON, KENTUCKY	.5415	.4468	174323.	94.	91	170	40	135
115	LIMA, OHIO	.6580	.2492	171472.	63.	148	10	39	28
116	LINCOLN, NEBRASKA	.6186	.3815	167972.	90.	129	129	36	124
117	LITTLE ROCK-NORTH LITTLE ROCK, ARK.	.4775	.3542	323296.	69.	69	109	100	44
118	LORAIN-ELYRIA, OHIO	.5279	.2676	256843.	67.	82	18	71	111
119	LOS ANGELES-LONG BEACH, CALIF.	.3879	.5145	7032075.	117.	36	184	168	175
120	LOUISVILLE, KY.-IND.	.5379	.3366	826553.	89.	90	92	151	75
121	LOWELL, MASS.	.7231	.3580	212860.	89.	170	113	55	171
122	LUBBOCK, TEXAS	.2677	.3946	179295.	79.	10	142	43	72
123	LYNCHBURG, VIRGINIA	.5976	.3163	123474.	59.	120	56	9	16
124	MACON, GEORGIA	.4333	.4139	206342.	61.	56	158	51	21
125	MADISON, WISCONSIN	.5799	.4364	290272.	123.	106	169	86	179
129	MEMPHIS, TENN.-ARK.	.4314	.4270	770120.	72.	54	164	149	54
131	MIAMI, FLORIDA	.2524	.4590	1267792.	129.	6	173	166	182
133	MILWAUKEE, WISCONSIN	.6210	.4018	1403688.	104.	131	149	172	161
134	MINNEAPOLIS-ST. PAUL, MINN.	.5903	.3479	1813647.	119.	115	101	175	177
135	MOBILE, ALABAMA	.3526	.3174	376690.	53.	25	60	113	4
137	MONROE, LA.	.4087	.3387	115387.	56.	45	93	2	8
138	MONTGOMERY, ALABAMA	.4739	.3861	201325.	55.	67	137	49	6
139	MUNCIE, INDIANA	.6059	.2944	129219.	77.	125	37	13	65
140	MUSKEGON-MUSKEGON HEIGHTS, MICH.	.5603	.2125	157426.	76.	100	3	30	64
142	NASHVILLE, TENNESSEE	.4766	.3775	541108.	80.	68	126	132	76
143	NEW BEDFORD, MASS.	.8085	.4652	152642.	61.	183	177	26	22
144	NEW BRITAIN, CONNECTICUT	.6008	.4244	145269.	90.	122	161	22	175
145	NEW HAVEN, CONNECTICUT	.6897	.4305	355538.	115.	162	166	109	170
146	NEW LONDON-GROTON-NORWICH, CONN.	.6359	.3783	208412.	58.	139	127	53	118
147	NEW ORLEANS, LA.	.5527	.4865	1045809.	78.	96	183	160	69
148	NEW YORK, NEW YORK	.7133	.6318	11571899.	115.	168	188	189	171

149	NEWARK, NEW JERSEY	.6874	.4665	1856556.	119.	161	178	176	178
150	NEWPORT NEWS-HAMPTON, VA.	.3374	.3967	292159.	89.	23	143	87	119
151	NORFOLK-PORTSMOUTH, VA.	.3815	.4506	680600.	84.	33	171	144	98
154	OGDEN, UTAH	.4886	.3043	126278.	80.	70	49	11	77
155	OKLAHOMA CITY, OKLAHOMA	.4389	.3230	640889.	85.	58	69	141	99
156	OMAHA, NEBR.-IOWA	.6494	.3669	540142.	92.	144	119	131	131
157	ORLANDO, FLORIDA	.2604	.3030	428003.	90.	7	43	122	126
160	PATERSON-CLIFTON-PASSAIC, N.J.	.5649	.3737	1358794.	125.	101	122	169	181
161	PENSACOLA, FLORIDA	.3125	.2921	243075.	71.	19	34	68	52
162	PEORIA, ILLINOIS	.5846	.2838	341979.	85.	110	26	104	100
164	PHILADELPHIA, PA.-N.J.	.6640	.3290	4817914.	97.	151	76	186	147
165	PHOENIX, ARIZONA	.2069	.3368	967522.	102.	4	87	157	152
167	PITTSBURGH, PENNSYLVANIA	.6689	.3223	2401245.	83.	154	68	181	92
169	PORTLAND, MAINE	.7604	.3942	141625.	81.	180	141	18	84
170	PORTLAND, ORE.-WASH.	.5489	.3503	1009129.	96.	94	103	158	143
171	PROVIDENCE-PAWTUCKET-WARWICK- R.I.-MAS	.7316	.4107	910781.	70.	172	154	155	49
172	PROVO-OREM, UTAH	.4515	.3295	137776.	81.	62	77	16	85
173	PUEBLO, COLORADO	.5559	.2716	118238.	63.	97	20	7	39
174	RACINE, WISCONSIN	.6493	.3040	170838.	94.	143	46	38	136
175	RALEIGH, N.C.	.4312	.4115	228453.	75.	53	155	60	61
176	READING, PENNSYLVANIA	.7548	.2798	296382.	69.	176	24	90	45
178	RICHMOND, VIRGINIA	.5012	.3865	518319.	87.	73	136	126	112
179	ROANOKE, VIRGINIA	.5363	.3039	181436.	74.	69	45	44	58
181	ROCHESTER, NEW YORK	.7112	.3323	882667.	109.	167	78	154	164
182	ROCKFORD, ILLINOIS	.5455	.3283	272063.	95.	93	75	77	139
183	SACRAMENTO, CALIFORNIA	.3055	.3636	800592.	103.	15	133	150	156
184	SAGINAW, MICHIGAN	.6268	.2275	219743.	69.	135	5	57	122
186	ST. LOUIS, MO.-ILL.	.6152	.3539	2363017.	86.	128	107	180	164
189	SALT LAKE CITY, UTAH	.4962	.3260	557635.	86.	72	72	134	105
191	SAN ANTONIO, TEXAS	.4254	.3607	864014.	79.	49	115	153	73
192	SAN BERNARDINO-RIVERSIDE-ONTARIO, CALI	.2665	.3622	1143146.	90.	9	116	163	127
193	SAN DIEGO, CALIFORNIA	.2822	.4359	1357854.	116.	14	168	168	172
194	SAN FRANCISCO-OAKLAND, CALIF.	.5349	.4436	3109519.	132.	87	182	184	183
195	SAN JOSE, CALIFORNIA	.2639	.3825	1064714.	143.	8	131	161	188
196	SANTA BARBARA, CALIFORNIA	.3899	.4621	264324.	113.	36	176	75	168
198	SAVANNAH, GEORGIA	.4727	.4259	187767.	64.	66	163	46	32
199	SCRANTON, PENNSYLVANIA	.9102	.3660	234107.	58.	188	113	65	15
200	SEATTLE-EVERETT, WASH.	.5024	.3514	1421869.	112.	75	104	173	167
202	SHREVEPORT, LA.	.4114	.3515	294703.	60.	46	105	88	18
203	SIOUX CITY, IOWA-NEB.	.7791	.3042	116189.	68.	179	48	4	40
205	SOUTH BEND, INDIANA	.5951	.2269	280031.	80.	117	6	79	78
206	SPOKANE, WASHINGTON	.5573	.3084	287487.	73.	93	52	84	56
207	SPRINGFIELD, ILLINOIS	.6702	.3352	161335.	83.	156	82	32	93
208	SPRINGFIELD, MO.	.5893	.3148	152929.	74.	107	54	27	59
209	SPRINGFIELD, OHIO	.6640	.3199	157115.	72.	150	64	29	55
210	SPRINGFEILD-CHICOPEE-HOLYOKE, MASS.-CON	.6817	.4067	529922.	80.	159	151	128	79
211	STAMFORD, CONNECTICUT	.5769	.3826	206419.	149.	105	132	52	189
212	STEUBENVILLE-WEIRTON, OHIO-W.VA.	.6541	.2577	165627.	56.	146	12	35	9
213	STOCKTON, CALIFORNIA	.4490	.3860	290208.	87.	60	136	85	88
214	SYRACUSE, NEW YORK	.6868	.3424	636507.	96.	160	95	140	144
215	TACOMA, WASHINGTON	.5340	.3397	411027.	86.	85	94	120	106
217	TAMPA-ST. PETERSBURG, FLORIDA	.3093	.2548	1012594.	86.	18	11	159	107
218	TERRE HAUTE, INDIANA	.7690	.2415	175143.	62.	177	8	42	24
220	TOLEDO, OHIO-MICH.	.6742	.2935	692571.	87.	157	36	145	113
221	TOPEKA, KANSAS	.5817	.3571	155322.	87.	106	112	28	114
222	TRENTON, NEW JERSEY	.6486	.3490	303968.	109.	142	102	94	165
223	TUCSON, ARIZONA	.2061	.3462	351667.	92.	3	99	108	132
224	TULSA, OKLAHOMA	.4927	.3171	476945.	85.	71	58	123	101

225	TUSCALOOSA, ALABAMA	.4156	.3924	116029.	56.	47	140	3	10
227	UTICA-ROME, NEW YORK	-.7488	.3367	340670.	69.	174	86	103	46
230	WACO, TEXAS	.4647	.3428	147553.	62.	64	96	24	25
231	WASHINGTON, D.C.-MD.-VA.	.3889	.5403	2861123.	137.	37	186	183	186
232	WATERBURY, CONNECTICUT	-.6512	.3823	208956.	82.	145	130	54	89
233	WATERLOO, IOWA	-.5709	.2662	132916.	82.	102	15	14	90
234	WEST PALM BEACH, FLORIDA	.2789	.3240	348753.	102.	12	70	107	153
235	WHEELING, W. V.-OHIO	-.8114	.3030	182712.	49.	184	42	45	3
236	WICHITA, KANSAS	.3815	.3528	389352.	83.	34	106	116	94
237	WICHITA FALLS, TEXAS	.4692	.3380	127621.	66.	65	90	12	35
238	WILKES-BARRE-HAZLETON, PA.	-.8882	.3343	342301.	56.	187	81	105	11
239	WILMINGTON, DEL.-N.J.-MD.	-.5058	.3169	499493.	91.	77	57	125	128
241	WORCESTER, MASSACHUSETTS	-.7300	.4015	344320.	79.	171	148	106	74
242	YORK, PENNSYLVANIA	-.6559	.2745	329540.	65.	147	22	102	33
243	YOUNGSTOWN-WARREN, OHIO	.5977	.2469	536003.	78.	121	9	129	70

4.2.2. Correlations

A second type of analysis that can be performed on the data is the calculation of correlations among the independent variables. Table 3 gives the simple correlation coefficients and Table 4 the Spearman rank correlation coefficients, calculated on tract data for the city of New Haven. These data are being used for model calibration, since at the time these analyses were being carried out, New Haven was the only available source of tract incidence data. A correlation coefficient of 1 indicates the two variables differ at most by a constant scale factor. (For example, the same measurements both in inches and in feet differ by the constant scale factor 12 inches/foot.) The interpretation of the simple correlation coefficient assumes the variables are normally distributed throughout the whole population, while for the rank correlation coefficient no assumption on the distribution of the variables is made. The first column of each table gives the correlation coefficients between each of the independent variables and the dependent variable.

The analysis of the correlations is used in the modelling process in two ways:

1) to avoid fitting error

2) to substitute for unavailable variables.

If two (or more) parameters are assumed to vary independently of each other whereas in fact they are correlated, spurious coefficients may occur in the process of curve fitting. If the resultant model is to be used simply for predictions based on the insertion of observed values of the model parameters, no real harm is done, because presumably the observed

	Number of EBL	Total Housing	Rental Units	Deteriorating Units	Dilapidated Units	Units Built Before 1940	Total Population	Children ≤ 4 Years	Children 5 Years	Children 6 Years	Median Income	Population Crowded (1.01)	Female Household Heads	Population Crowded (1.51)	Median Rent	% Built Before 1940
1. Number of EBL	-															
2. Total Housing	.2943	-														
3. Rental Units	.3578	.8891	-													
4. Deteriorating Units	.7224	.3931	.5460	-												
5. Dilapidated Units	.3768	.2815	.5017	.6061	-											
6. Units Built Before 1940	.4484	.7271	.8479	.6967	.5609	-										
7. Total Population	.4129	.8381	.6539	.3407	.1271	.4694	-									
8. Children ≤ 4 Years	.6478	.6258	.4729	.4629	.2019	.3841	.8307	-								
9. Children ≤ 5 Years	.6831	.6302	.5056	.5614	.3137	.4386	.8201	.9704	-							
10. Children ≤ 6 Years	.6658	.6523	.5229	.5468	.3360	.4485	.8303	.9446	.9814	-						
11. Female Household Heads	.6386	.6017	.6392	.6777	.6027	.6111	.6740	.7836	.8517	.8699	-					
12. Median Income	-.3053	.0498	-.2494	-.4940	-.6010	-.3463	.0126	-.1492	-.2113	-.1992	-.4086	-				
13. Population Crowded (1.01)	.7441	.5386	.5074	.6295	.3648	.4514	.7332	.9135	.9445	.9311	.8530	-.3916	-			
14. Population Crowded (1.51)	.8483	.4985	.4982	.7206	.3774	.5196	.6533	.8255	.8623	.8476	.7834	-.3676	.9469	-		
15. Median Rent	-.3141	.1926	.0367	-.4164	-.3197	-.2434	.1175	-.1254	-.1911	-.1439	-.1905	.7368	-.3379	-.3754	-	
16. % Built Before 1940	-.2116	.0484	-.0903	-.1487	-.1042	.5608	-.1455	-.1017	-.0840	-.0862	.0684	-.5149	.0136	.1697	-.6006	-

Note that a negative coefficient denotes a correlation between a variable and the reciprocal of another. It is the magnitude coefficient that measures the "strength" of the correlation.

Table 3: Simple Correlation Coefficients

	Number of EBL	Total Housing	Rental Units	Deteriorating Units	Dilapidated Units	Units Built Before 1940	Total Population	Children ≤ 4 Years	Children 5 Years	Children 6 Years	Female Household Heads	Median Income	Population Crowded (1.01)	Population Crowded (1.51)	Median Rent	% Built Before 1940
1. Number of EBL	-															
2. Total Housing	.2231	-														
3. Rental Units	.4158	.8184	-													
4. Deteriorating Units	.5599	.2605	.4234	-												
5. Dilapidated Units	.5255	.1450	.3229	.7353	-											
6. Units Built Before 1940	.4556	.6158	.7669	.6990	.5190	-										
7. Total Population	.3187	.8194	.6298	.2244	.0411	.3525	-									
8. Children ≤ 4 Years	.6009	.6535	.5507	.2671	.1759	.3437	.7488	-								
9. Children ≤ 5 Years	.7242	.6391	.5434	.3805	.2668	.3961	.7757	.9517	-							
10. Children ≤ 6 Years	.6793	.6325	.5250	.3386	.2069	.3955	.7913	.8860	.9511	-						
11. Female Household Heads	.7690	.5981	.6992	.5351	.4648	.6262	.6386	.7335	.8223	.7881	-					
12. Median Income	-.4821	.2091	-.0684	-.5539	-.5958	-.1949	.0586	.0640	-.0802	-.0498	-.2737	-				
13. Population Crowded (1.01)	.8398	.5151	.5822	.4893	.4223	.4532	.6552	.8615	.9224	.8887	.8786	-.3213	-			
14. Population Crowded (1.51)	.8630	.4844	.5663	.5972	.4637	.5057	.5835	.7559	.8496	.8165	.8567	-.3569	.8551	-		
15. Median Rent	-.4833	.2468	.1209	-.4966	-.4975	-.2104	.1761	-.0657	-.1664	-.0959	-.1650	.7146	-.3153	-.3736	-	
16. % Built Before 1940	.1812	-.2837	-.0604	.4902	.3155	.3275	-.3730	-.3822	-.2736	-.2442	-.1956	-.4902	-.1052	.0249	-.6200	-

Table 4: Spearman Rank Correlation Coefficients

values of the correlated variables will be correctly noted. If, however, the model is used for experimental simulation, e.g., varying selected parameters while leaving others fixed, serious distortions will ensue, very likely invalidating the experiment. A clarifying example follows:

Suppose we estimate that some variable Z is related to variables X and Y,

$$Z = aX^bY^c,$$

where we think that X and Y are independent of each other. Suppose, then, that the formulation is correct except that in fact, X and Y obey the relationship

$$Y = gX ,$$

so that in reality, the relationship between the dependent variable (Z) and the independent variable X is

$$Z = a'X^{b'}.$$

Therefore, formally

$$Z = aX^bY^c = aX^b(gX)^c = ag^cX^{b+c}.$$

Thus a calibration process for determining the numerical values of a, b and c may properly yield any values as long as they satisfy

$$ag^c = a'$$

$$b + c = b',$$

which has infinitely many solutions for fixed a', b', g. It is easy to see what varying X without changing Y (after a, b, and c have been calculated) will lead to incorrect values for Z.

Correlations among observations need not always give cause for alarm. They can be utilized fruitfully by allowing us to choose a variable when there are several available with differing precision of observation or when a desirable variable is unavailable. This is indeed true of the lead paint poisoning model. Values of several census descriptors believed to be directly related to the incidence of elevated blood lead levels are not included in the Census First Count tabulation, and are therefore unavailable for a preliminary model. Substitute parameters drawn from the available data were selected by examining the correlation coefficients in Tables 3 and 4. As an example, the incidence of lead poisoning is higher among the poor, but the income data are not yet available. Therefore, rent has been used as a substitute for income in the model. This process is of course, something of a gamble. The example at the beginning of this discussion used a pure linear correlation between variables. Nothing descriptive of human populations in an uncontrolled small sample can be expected to yield completely unambiguous results. We felt reasonably justified in making substitutions of variables where both correlation coefficients for a pair were greater than .7. After all, the substitution of one variable for another was to be made before a calibration, which means that the curve fitting program is determining coefficients and/or exponents for the substitute variables, and thus it is less likely to generate distortions than it would if the substitution of variables occurred after the model calibration.

4.2.3. Data Gaps

In addition to using the correlations to fill in gaps in the available data items, some extra modelling has been necessary. The most serious data gap is the lack of 1970 data on age and condition of housing.

To remedy the discontinuation in 1970, of tabulation of "condition of housing" data the Census Bureau has developed a method for predicting that information from the 1970 data and 1960 Fourth Count Census items including (1) units with central heating facilities, (2) rent, (3) units with crowded population; 1.01 per room, (4) multiple unit structures, (5) educational level of household head, (6) race of household head, (7) owned or rented, and (8) vacant units. The dwelling units are divided into categories determined by cutoff levels of the above attributes, and separate growth patterns are applied to each of the categories. However, since the Census Fourth Count data are not yet available, this model could not be used to furnish inputs to the lead poisoning models at the present stage.

The following make shift housing adjustment therefore was applied to the census tract level data, on the assumption that the results, if crude were likely to be more consistent with available 1970 EBL incidence data from New Haven tracts, than raw 1960 condition of housing figures.

Let: H_{60} - total 1960 housing units

H_{70} - total 1970 housing units

D_{60} - total 1960 unsound units

D_{70} - total 1970 unsound units

If $H_{70} \geq H_{60}$, then

$$D_{70} = D_{60}.$$

If $H_{70} < H_{60}$, then

$$D_{70} = \max\ [.05D_{60}, (D_{60} - .9(H_{60} - H_{70}))].$$

This means that, if the number of housing units in a tract has increased, we don't attempt to adjust the number of unsound units. If housing has decreased, it was assumed that 90 percent of the decrease represents demolition of unsound housing. However, if "net demolitions" appear to exceed the total 1960 unsound housing stock, then 5 percent of this unsound stock was assumed still standing. The choices of the .05 and .9 are arbitrary but have been chosen to approximate patterns stated to exist in New Haven.* (This "model" could not be applied usefully at the SMSA level because it would predict no change in unsound stock since almost all metropolitan areas have grown in the last 10 years. Units demolished in the central cities are more than counter-balanced by ones built in the suburbs.) Thus the housing figures used in the modelling process are not as current as one would desire and may be altered in future modelling procedures.

In response to a request from the Department of Housing and Urban Development, the present study includes a method for rough estimation of housing units in each SMSA believed to contain sufficient concentrations of lead on painted surfaces to constitute lead poisoning hazards. (Although this estimated number of "hazardous environments" has not yet been incorporated as a variable into the EBL models tested, the formulation of the method is presented here to make the current record complete.)

The housing estimation embodies the following assumptions:

(1) Almost all housing units constructed before 1940 in urbanized regions contain lead paint from original surfacing and/or

---

*Reported orally during field trip to office at New Haven Lead Poisoning Prevention Program.

subsequent refurbishment.

(2) About half of the housing units built between 1940 and 1949 in urbanized areas are similarly contaminated. (This is an informal estimate based on the reduction in the quantity of lead paint manufactured in this decade, relative to housing construction in the same period.)*

(3) Only 5% of housing constructed between 1950 and 1959 has interior lead paint, and post 1960 housing can be ignored in the estimates.

(4) Areas in which non-negligible fractions of the housing stock were largely unpainted or merely whitewashed (substantially in the rural South) or other areas in which the climate allowed painting materials to be selected without great concern for weathering properties, such as the south west and far west, can be distinguished crudely from high hazard concentration areas by differences in the fraction of dwelling units which are single unit structures.

Reasoning from (4) above, the estimates by age in (1), (2), and (3) are attenuated in the model by the fourth root of the fraction of dwelling units in multi-unit structures. (Weighting by the <u>actual</u> fractions of multi-unit structures leads to counts of contaminated dwelling units which are suspiciously low, so they have been adjusted upward by using the fourth root, the square root of the square root, of these fractions. The fourth root has desirable properties of being easy to calculate, and

---

*<u>Census of Manufactures</u> and <u>Minerals Yearbook</u> were analyzed to calculate these figures.

remaining smaller than 1.0 for fractions. It also compresses the range of the weights.) Typical values of these fractions in southern cities and SMSA's* are .3 and .2, respectively, in northeastern cities and SMSA's .7 and .5 respectively. The fourth roots of these numbers are, in the order listed, .74, .67, .91, and .84.

The count of housing units by age in each SMSA has to be estimated, because while 1970 national figures for % of housing by decade of construction have been published by the Bureau of the Census, counts by SMSA are not yet available. The estimates are made by multiplying the 1960 census counts by the ratios of national percentages in 1960 and 1970. Thus the estimate has the form

$$HL = \sqrt[4]{M}\left(H_{40/60}\,\frac{h_{40/70}}{h_{40/60}} + .5H_{50/60}\,\frac{h_{50/70}}{h_{50/60}} + .05H_{60/60}\,\frac{h_{60/70}}{h_{60/60}}\right)$$

where HL is the estimated number of contaminated housing units in an SMSA

$H_{40/60}$, $H_{50/60}$, $H_{60/60}$ are the counts of 1960 housing stock in the SMSA built before 1940, during 1940-49, and during 1950-59, according to 1960 census data.

$h_{40/60}$, $h_{50/60}$, $h_{60/60}$ are the fractions of the 1960 national housing stock built before 1940, during 1940-49, and during 1950-59, according to 1960 census data

$h_{40/70}$, $h_{50/70}$, $h_{60/70}$ are the fractions of the 1970 national housing stock built before 1940, during 1940-49 and during 1950-59, according to 1970 census data.

---

*City and County Data Book 1967, U.S. Department of Commerce, Bureau of the Census.

It should be noted that the estimates of EBL in children presuppose at least surface deterioration in the bulk of aging housing units, making paint with significant lead content easily accessible to the risk population of children 6 years of age or younger. The estimates of the above housing model, therefore, would identify only potential sources of lead poisoning.

Three obiter dicta are in order here: first, areas of potential hazard of the type described become active hazards rather shortly in periods of economic change and rapid migration, under the kind of dereliction of maintenance that usually accompanies such sociological phenomena, so that in effect, tracts of latent contaminated housing are apt to lose their latency virtually en masse.

Second, we, and others, have tacitly assumed that housing units dating from the period of pervasive use of lead paint, but not deteriorating or dilapidated according to census criteria, presented no immediate hazard. In fact, however, housing units with (possibly extensive) paint peeling from wall surfaces could be judged sound by census criteria if no other defects were present, because peeling paint can be repaired by simple maintenance effort.

Third, all housing conditions and socio-economic factors aside, any curious very young child, unless under physical restraint or under literally continuous supervision, can readily fracture the integument of a wall surface and ingest sufficient quantities of lead, (if present) to result in lead pathologies.

While these observations don't pertain to model construction, we offer them as insights gleaned from the analytical activities in the model development, which advert to an underlying motivation for the models: assessment of the lead paint hazard.

5. THE MODEL FORM

The choice of an initial model structure in the absence of a highly detailed explanatory hypothesis rests on two principal criteria: simplicity and conceptual plausibility. The first, which is more or less objective, affects ease of calibration and amenability to modification when inevitable improvements in data availability occur. (It is the nature of the universe that the data base for a first model effort is invariably so bad as to elicit desperate thoughts of total infeasibility.) Conceptual plausibility is somewhat subjective and is really a measure of how comfortable one is with whatever ideas support a given simple formulation.

The two simplest kinds of representation of the dependancy of a numerical measure of some phenomenon on the values of a set of others are grouping them additively (linear models) or multiplicatively (exponential models).

A linear model with 5 independent variables is of the form

$$(1)\quad Y = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + a_4X_4 + a_5X_5$$

A multiplicative model in 5 variables would be

$$(2)\quad Y = a_0 \cdot X_1^{a_1} \cdot X_2^{a_2} \cdot X_3^{a_3} \cdot X_4^{a_4} \cdot X_5^{a_5}$$

Notice that each of these has 6 formal parameters; a constant term (which defines an initial level in the additive model and which serves as a scaling factor in the multiplicative model) and a parameter for each variable, which are additive coefficients in the linear model or multiplicative exponents in the exponential model. Notice also in the multiplicative version, that unless a_0 is zero (in which case the model is just the constant, Y = 0), it is not possible to represent zero values of the dependent variable, which would require one of the exponents to be $-\infty$ (leading again to the constant model, Y = 0).

Barring the feasibility of a child-by-child enumeration by blood levels estimated along the lines of the very detailed physiological model of Batelle Institute,[40] no attempt at formulation of a linear model was remotely successful.

The notion of rates of incidence in the population as conditional probabilities, i.e., defining incidences as the intersection or concomitance of various conditions believed to be factors in ingestion of lead by children, leads naturally to a formulation multiplying relative frequencies of occurrence of suspected causative factors. For example, one might think of the expected number of elevated blood lead levels as the child population times the probability that a child will live in a house offering a hazard, times the probability that the child will ingest the hazardous material. Of course, this last probability is not known per se, but factors believed relating to it are available from the data base.

The simplest multiplicative model would apply if the concurrent presence of the listed factors <u>insured</u> the presence of lead poisoning. However, the data base and our understanding of the lead poisoning process do not lend themselves to such an exact interpretation. Exponents therefore, are introduced to represent the less sure situation and to reflect the different levels of contributions of the factors. That is, the "probability" factors are exponentiated yielding additional degrees of freedom. Thus, instead of an estimating expression for a quantity E of the form $E = p_1 p_2 \cdots p_k$, which is fixed, given the values of factors p_i, we would have $E = p_1^{a_1} p_2^{a_2} \cdots p_k^{a_k}$ where the values of the exponents a_i can be varied to obtain values for E consistent with observations. These exponents also may help overcome difficulties occurring because of dependencies between the supposedly independent variables.

5.1. Curve Fitting for Parameter Estimation - Log Linearity

In the models described below several parameters must be estimated; this procedure is called curve fitting. The standard procedure involves the "least squares estimate of a straight line." However, the model form chosen for the lead paint poisoning model is multiplicative rather than linear.

While the equation $Y = aX + b$ with a and b as parameters represents only straight lines, the equation $Y = bX^a$ can represent a mathematically simple, but wide ranging, family of curves. This exponential form "induces" additional degrees of freedom for attempts to fit simple curves to observed data points and still <u>permits one to use linear curve fitting methods by taking logs of both sides of the equation</u>. Thus if

(1) $Y = bX^a$ then

(2) $\log Y = a \log X + \log b$,

and changing variables just for neatness: letting $Z = \log Y$, $W = \log X$, and $b' = \log b$, we have

(3) $Z = aW + b'$. Similarily for many variables, the equation

(1)' $Y = a_0 X_1^{a_1} X_2^{a_2} \ldots X_k^{a_k}$ is equivalent to

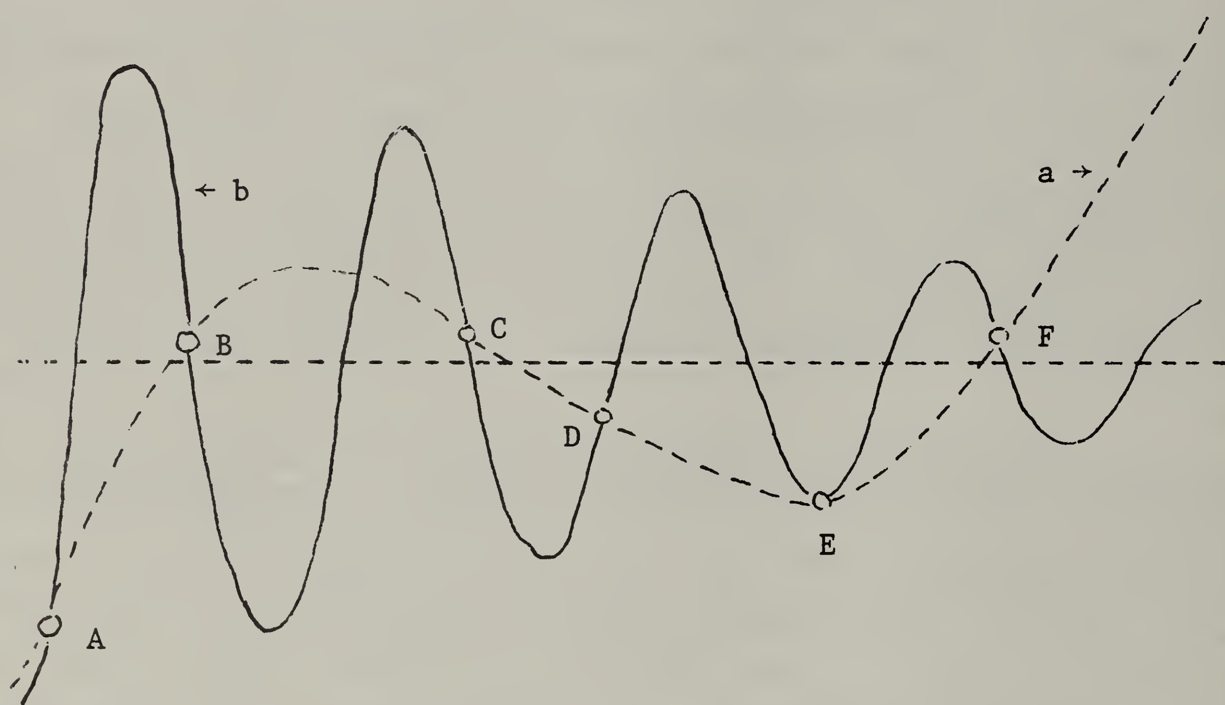
(3)' $Z = a'_0 + a_1 W_1 + a_2 W_2 + \ldots a_k W_k$.

The name "log linear" is used to describe multiplicative relationships of form (1), (1)' for reasons which are obvious from (3), (3)'. There is one special danger (a psychological flaw) in using models of this type, in that it is easy to forget that the linear form used in the calculations must be exponentiated to interpret results. Let us clarify this with an example.

Supposing we have deduced the model $Y = 1.6X + 2$ from observations of some physical phenomenon. Now, if e.g., for $X = 2$ the true value of Y is 5, then the relative error is $\frac{1.6 \cdot 2 + 2 - 5}{5} = \frac{5.2 - 5}{5} = \frac{.2}{5}$ or .04 if X and Y are the variables of interest. If, however, they are the <u>logarithms</u> of these variables then our error is the relative error in the logarithms and the actual relative error is $\frac{e^{5.2} - e^5}{e^5} = .22$, more than 5 times as much.

5.2. <u>Judging the Quality of a Model</u>

A common practice in judging the validity of a model is to calculate on the original data set, the average difference in magnitude between predicted and observed values of the variable to be estimated, divided by the value of a typical observation, that is, the prediction error as a percentage of the observed value. If this figure is small, the model is asserted to be a good one. In fact it may be nothing of the kind. The number described is an excellent measure of how well the formulation fit a given set of observations in question but purely by luck the fit <u>could</u> be perfect while the underlying hypothesis by which the functional form was articulated is dead wrong, because a given set of points can lie on more than one curve. The figure below illustrates this



In the figure the points A through F lie on the broken line <u>a</u> and on the solid curve <u>b</u>. If we were predicting behavior based on the fit to curve <u>a</u> of a phenomenon actually governed by the curve <u>b</u>, we would certainly have more and more to regret the further we moved to the right because <u>b</u> is presumably converging to the value on the level dotted line while <u>a</u> is growing.*

---

*Note that in general, k + 1 points can be fitted exactly with a polynomial curve of degree k.

Therefore, judging the quality of a model, requires, in addition to mechanical criteria used in curve fitting, some determination of whether the model actually mirrors the process it is supposed to describe. Referring back for a moment to the determination of "goodness of fit", Appendix A consists of the output from a computer run of a standard curve fitting program. The printout is from a calibration of one of the versions of the EBL incidence model. ("Calibration" is the calculation of values for the parameters of the model.) The statistical measure of accuracy of the fit is a slight generalization of the one outlined in the example. It is the square of the multiple correlation coefficient;* it is denoted by the symbol R^2, (usually read literally: "capital R square") and has the formulation:

$$R^2 = 1 - \frac{\Sigma(y_i - \hat{y}_i)^2}{\Sigma\, y_i^2}$$

Thus it is one (1.0) minus the ratio of the sum of the squares of the differences between the observed values of the dependent variable and their predicted values, and to the sum of the squares of the observed values, or using the terminology of the printout, 1.0 minus "the residual sum of squares divided by the total sum of squares."

---

*Natrella, M. G., Experimental Statistics, National Bureau of Standards Handbook 91, Washington, D.C., August 1, 1963.

The criteria we have used for acceptability of fit in the first set of experimental model formulations are:

For a given model, whether or not R^2 exceeds .9

For refinements of acceptable models, produced by the introduction of additional variables, whether or not R^2 for the refined model is increased by .05. (This sets a tradeoff rate for complexity vs. fit improvement.)

As we have implied previously, "fitness" of a model involves more than the "quality of fit". One relatively simple test that can be applied individually to the calculated parameters (coefficients and exponents) is whether their signs are consistent with our judgments (or prejudices). For example, we would expect the variable "family income" to appear here in a multiplicative model, with a negative exponent, because we believe the incidence of EBL in an area to decrease as the income level rises. Enumerating a few more examples, we expect positive exponents for

Crowding (a concomitant of bad housing and poverty)

Proportion of "bad" housing

Population of young children

Population

Number of female household heads (implying straitened economic circumstances and reduced capability for supervision of young children)

If the characteristics of a parameter are "counterintuitive" the data sample may be distorted, the general formulation of the model may be inappropriate, the variable may have been selected by erroneous reasoning, or indeed, the rationale for the sign criterion for the parameter might be mistaken. One immediate example of possible ambiguity occurs in considering the "crowding" variables. Early in the study, during initial selection of relevant variables, several staff members put forth the idea that extreme crowding might militate against lead ingestion, because it would tend to afford a paint eater less opportunity to remove chips from walls undetected.

In general, in the models that follow, wherever we state that a calculated parameter is "wrong" or "incorrect" it is to be interpreted in the sense described above.

One final comment. The "Model" implied by the title of this paper is at this stage a composite of models 1 and 2.9 below, which was employed to produce sample estimates. It should not be assumed that we consider this model output as authoritative in any way. It is an interim model from a study in progress and should be regarded that way. If the reader requires a further caveat, recall section 5.1 and consider that the values of R^2 for the models below pertain to calibration runs with the models entered in "log linear" form and that the true measures of fit would be somewhat less impressive.

6. THE MODELS

6.1. The Variables and Symbols Used in the Models

$a_0, a_1, \ldots$	parameters estimated by the curve fitting process
C_{101}	crowded population ($\geq$ 1.01 per room)
C_{151}	crowded population ($\geq$ 1.51 per room)
C_F	crowded(1.01) population with female household head
D	dilapidated or deteriorating housing units (number)
E	number of children with elevated blood lead levels ($\geq$ 40 μg/100 ml.)
F	number of female household heads
H	number of housing units
I	fraction of child population with elevated blood lead levels
K_4	number of children 4 years old or younger
K_6	number of children 6 years old or younger
L	number of housing units lacking one or more plumbing facilities
P	total population
R	median rent \$/mo.
R_{40}	number of units with rent $\leq$ \$40
R_{60}	number of units with rent $\leq$ \$60
Y	median family income \$/ann.

6.2. Model 1

$$E = K_6 \left(\frac{D}{H}\right) a_0$$

where $a_0 = .24$.

In this model a_0 was estimated by averaging known incidence rates from existing programs.* These programs have for the most part screened children living in neighborhoods with poor housing. If one makes the assumption that the number of children per dwelling unit is the same for poor housing as for all housing, then $K_6\left(\frac{D}{H}\right)$ is the number of children 6 years old or less who live in poor housing.

6.3. Model 2

The versions of Model 2 are exponential in form and all use unsound housing as a variable.

6.3.1. Model 2.1

$$E = a_0\left(\frac{F}{P}\right)^{a_1}\left(\frac{D}{H}\right)^{a_2}\left(\frac{K_4}{P}\right)^{a_3}\left(\frac{C_{151}}{H}\right)^{a_4}$$

where $a_0 = e^{-2.222}$

$a_1 = .3197$

$a_2 = .1305$

$a_3 = -.7846$

$a_4 = 1.757$

Residual sum of squares 7.37

Total sum of squares 26.24 $R^2 = .719$

The fit of this model is not too good and a_3 has the wrong sign, since one would expect areas with more children to have a higher incidence of lead poisoning.

---

*See Table 4 of "National Estimates of Lead Based Paint Poisoning of Children (Estimated by Standard Metropolitan Statistical Area)"

6.3.2. Model 2.2

$$E = a_0 \left(\frac{F}{P}\right)^{a_1} \left(\frac{D}{H}\right)^{a_2} \left(\frac{K_4}{P}\right)^{a_3} (C_{151})^{a_4}$$

where $a_0 = e^{-3.912}$

$a_1 = -.1911$

$a_2 = .3232$

$a_3 = .1969$

$a_4 = 1.052$

Residual sum of squares 5.53

Total sum of squares 122.75 $R^2 = .955$

Although the fit for this model is very good, the sign of a_1 is wrong.

6.3.3. Model 2.3

$$E = a_0 \left(\frac{F}{P}\right)^{a_1} \left(\frac{D}{H}\right)^{a_2} (Y)^{a_3} \left(\frac{K_4}{P}\right)^{a_4} (C_{151})^{a_5}$$

where $a_0 = e^{12.798}$

$a_1 = -.6470$

$a_2 = .3483$

$a_3 = -2.1726$

$a_4 = .4338$

$a_5 = .9413$

Residual sum of squares 4.93

Total sum of squares 122.75 $R^2 = .960$

Although the fit is quite good, the sign of a_1, is incorrect.

6.3.4. Model 2.4

$$I = a_0 \left(\frac{F}{P}\right)^{a_1} \left(\frac{D}{H}\right)^{a_2} (Y)^{a_3} \left(\frac{K_4}{P}\right)^{a_4} (C_{151})^{a_5}$$

where $a_0 = e^{6.895}$

$a_1 = -.4757$

$a_2 = .3560$

$a_3 = -.9926$

$a_4 = .2663$

$a_5 = -.3433$

Residual sum of squares 1.59

Total sum of squares 40.2 $R^2 = .960$

Although the fit is quite good, the signs of both a_1 and a_5 are wrong.

6.3.5. Model 2.5

$$I = a_0 \left(\frac{K_6}{P}\right)^{a_1} \left(\frac{F}{P}\right)^{a_2} \left(\frac{D}{H}\right)^{a_3} (Y)^{a_4} \left(\frac{C_{101}}{P}\right)^{a_5}$$

where $a_0 = e^{7.654}$

$a_1 = .9697$

$a_2 = -.5163$

$a_3 = .3113$

$a_4 = -.9008$

$a_5 = -.374$

Residual sum of squares 1.34

Total sum of squares 40.2 $R^2 = .967$

Although the fit is quite good, the signs of a_1 and a_5 are wrong.

6.3.6. Model 2.6

$$I = a_0 \left(\frac{K_6}{P}\right)^{a_1} \left(\frac{F}{P}\right)^{a_2} \left(\frac{D}{H}\right)^{a_3} (R)^{a_4} \left(\frac{C_{101}}{P}\right)^{a_5}$$

where $a_0 = e^{2.4785}$

$a_1 = .8806$

$a_2 = -.3186$

$a_3 = .3062$

$a_4 = -.6207$

$a_5 = -.6908$

Residual sum of squares 1.32

Total sum of squares 40.2 $R^2 = .967$

Although the fit is quite good, the signs of a_2 and a_5 are wrong.

6.3.7. Model 2.7

$$I = a_0 \left(\frac{K_6}{P}\right)^{a_1} \left(\frac{F}{P}\right)^{a_2} \left(\frac{D}{H}\right)^{a_3} R^{a_4}$$

where $a_0 = e^{.5306}$

$a_1 = .6416$

$a_2 = -.4574$

$a_3 = .3001$

$a_4 = .1729$

Residual sum of squares 1.5

Total sum of squares 40.2 $R^2 = .963$

Although the fit is quite good, the signs of a_2 and a_4 are wrong.

6.3.8. Model 2.8

$$I = a_0 \left(\frac{K_6}{P}\right)^{a_1} \left(\frac{D}{H}\right)^{a_2} (R)^{a_3}$$

where $a_0 = e^{-.3362}$

$a_1 = .3003$

$a_2 = .2485$

$a_3 = .0129$

Residual sum of squares 2.21

Total sum of squares 40.2 $R^2 = .945$

Although the fit is quite good, the sign of a_3 is wrong.

6.3.9. Model 2.9

$$I = a_0 \left(\frac{K_6}{P}\right)^{a_1} \left(\frac{D}{H}\right)^{a_2}$$

where $a_0 = e^{-.2914}$

$a_1 = .2967$

$a_2 = .2484$

Residual sum of squares 2.21

Total sum of squares 40.2 $R^2 = .945$

The fit is good, and the signs of all parameters are correct. This is the "best" Model 2 version.

6.4. Model 3

All of the versions of Models 1 and 2 involved the poor housing variable. However, because this will not be available for 1970, an attempt was made to construct a model using only currently available 1970 Census First Count Data items.

6.4.1. Model 3.1

$$I = a_0 \left(\frac{K_6}{P}\right)^{a_1} (L)^{a_2}$$

where $a_0 = e^{.8428}$

$a_1 = .6747$

$a_2 = .2479$

Residual sum of squares 3.71

Total sum of squares 40.20 $R^2 = .908$

The fit is good and the signs of all parameters are correct.

6.4.2. Model 3.2

$$I = a_0\left(\frac{K_6}{P}\right)^{a_1} (L)^{a_2}\left(\frac{C_F}{P}\right)^{a_3}$$

where $a_0 = e^{.8853}$

$a_1 = .5825$

$a_2 = .0712$

$a_3 = .2354$

Residual sum of squares 3.68

Total sum of squares 40.2 $R^2 = .908$

The fit is good and all signs of parameters are correct. However the third factor only reduces the residual sum of squares .03 from model 3.1, so that is the preferred model.

6.4.3. Model 3.3

$$I = a_0\left(\frac{K_6}{P}\right)^{a_1} (L)^{a_2}\left(\frac{C_F}{P}\right)^{a_3} (R_{40})^{a_4}$$

where $a_0 = e^{1.9422}$

$a_1 = .6104$

$a_2 = .4743$

$a_3 = .3481$

$a_4 = -.3603$

Residual sum of squares 3.18

Total sum of squares 40.2 $R^2 = .921$

Although the residual sum of squares has been reduced from Model 3.2, the sign of a_4 is wrong.

6.4.4. Model 3.4

$$I = a_0 \left(\frac{K_6}{P}\right)^{a_1} (L)^{a_2} \left(\frac{C_F}{P}\right)^{a_3} (R_{60})^{a_4}$$

where $a_0 = e^{1.6604}$

$a_1 = .6869$

$a_2 = .3898$

$a_3 = .1670$

$a_4 = -.2349$

Residual sum of squares .32

Total sum of squares 4.78 $R^2 = .933$

This fit was done as a weighted regression where the weights were the fraction of total child population of each census tract screened. The sign of a_4 is incorrect.

7. FUTURE MODELLING WORK

Future modelling work can be divided into four areas:

a. further variations similar to those described above

b. modelling of the dilapidated and deteriorating housing for 1970

c. correction for the bias in the New Haven data

d. modelling poisoning in areas of less high risk.

Several other variations of models similar to those described above can be tried. However, it should be noted that in all the models tried so far the child population and the extent of unsound housing have been the strongest factors in the reduction of the sum of squared residuals. All other variables have contributed significantly less. Therefore, although in the interests of completeness other variations should be tried, little improvement is to be expected of this effort.

Since the quality of housing is apparently a major determinant of lead poisoning, it is hoped that obtaining a model for more current unsound housing figures will improve the simple version of Models 1 and 2. A procedure for this improvement was described earlier in this report but has not yet been implemented.

The rates obtained from the New Haven data may have led to too high a national projection. This is believed to be because screening in New Haven has been only in very high risk areas. A possible way to overcome the difficulty is to apply the rate obtained from the model only to high risk children (children in bad housing, e.g.) rather than to all children under 6 years old. This procedure has been used in the current effort. Another possibility is to consider only children in the central city instead of all children in the SMSA. If there were data on lead poisoning of children from middle income families, then one might be able to calculate different rates for high risk and lower risk children. The lack of good incidence data is the major difficulty in the modelling process. Chapter 4 of "Data Collection and Assimilation for the Lead Paint Poisoning Model is devoted entirely to a discussion of this problem." As the data from Aurora, Illinois, mentioned in that report, become available this will be alleviated somewhat. However, if these data are used for model development as it seems it will be necessary to do, another source must be found for model validation.

APPENDIX A

Sample Calibration Output

APPENDIX A

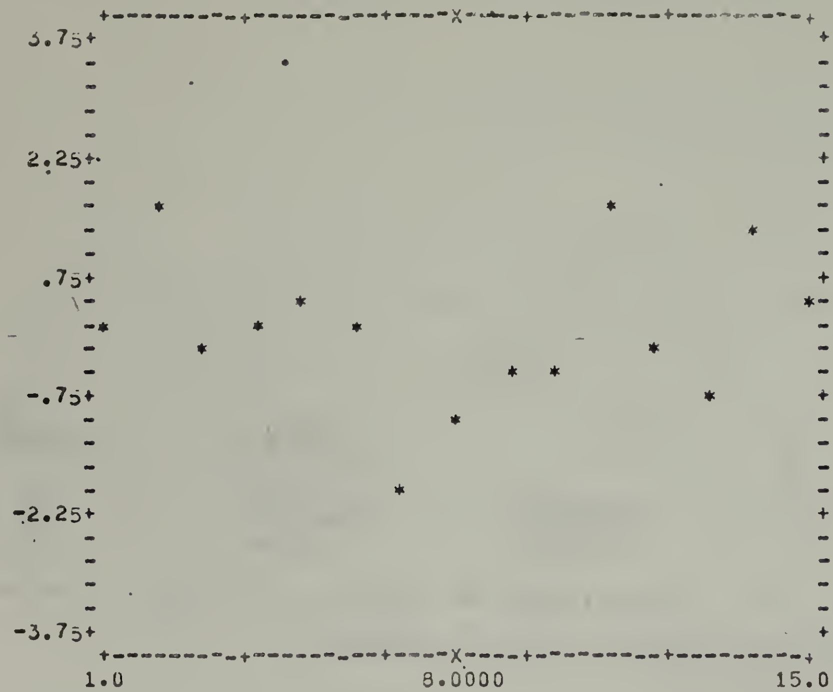
OMNITAB NEW HAVEN CENSUS DATA

LEAST SQUARES FIT FOR DATA IN COLUMN 32
AS A LINEAR FUNCTION OF 3 PREDICTOR VARIABLES IN COLUMNS 33, 34, 36
USING 15 NON-ZERO WEIGHTS = 1.0000000

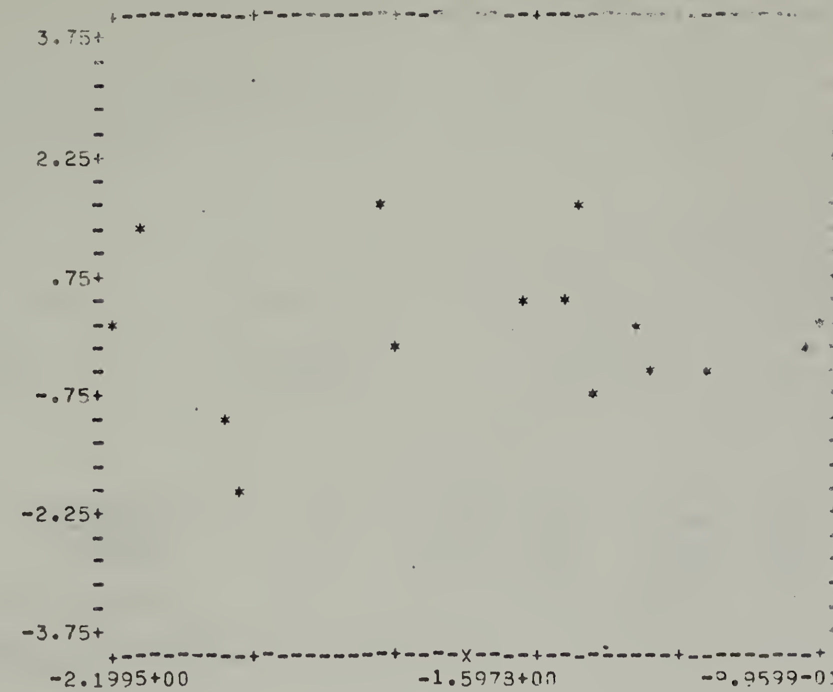
ROW	PREDICTOR VARIABLES IN COL. 34	COL. 36	DATA COL. 32	PREDICTED VALUES	STD. DEV. OF PRED. VALUES	RESIDUALS	STD. RES.	WEIGHTS
1	-2.12182	-1.54642	-1.2367626	-1.3051603	.14843157	.068397649	.17	1.000
2	-2.01433	-3.42361	-1.0185696	-1.7396428	.12806924	.72107319	1.76	1.000
3	-1.75201	-.761874	-1.0775589	-1.0034817	.19908509	-.074077188	-.19	1.000
4	-1.66438	-.848304	-.96742009	-.99598669	.21431132	.028566605	.08	1.000
5	-2.86152	-1.43077	-1.4069137	-1.4959118	.37987438	.088998158	.44	1.000
6	-2.27201	-4.96688	-2.1972246	-2.1995166	.22227629	.0022920437	.01	1.000
7	-2.07063	-4.28063	-2.7300291	-1.9692692	.16958504	-.76075988	-1.93	1.000
8	-2.25221	-4.18523	-2.4277462	-1.9994434	.17756684	-.42830484	-1.10	1.000
9	-1.89577	-1.30305	-1.4087672	-1.1776216	.15482085	-.23114563	-.58	1.000
10	-1.77609	-1.84264	-1.5141277	-1.2761707	.15533849	-.23795703	-.59	1.000
11	-1.98925	-2.02887	-.74721441	-1.3856878	.12032645	.63847335	1.55	1.000
12	-1.76185	-3.59498	-1.7917595	-1.7073039	.18020458	-.084455612	-.22	1.000
13	-2.05004	-1.85715	-1.7206636	-1.3610603	.12729977	-.35960326	-.88	1.000
14	-1.98450	-5.05303	-1.6468254	-2.1356239	.22420901	.48879843	1.33	1.000
15	-1.94773	-2.24700	-1.2878543	-1.4275584	.11783858	.13970410	.34	1.000

OMNITAB NEW HAVEN CENSUS DATA

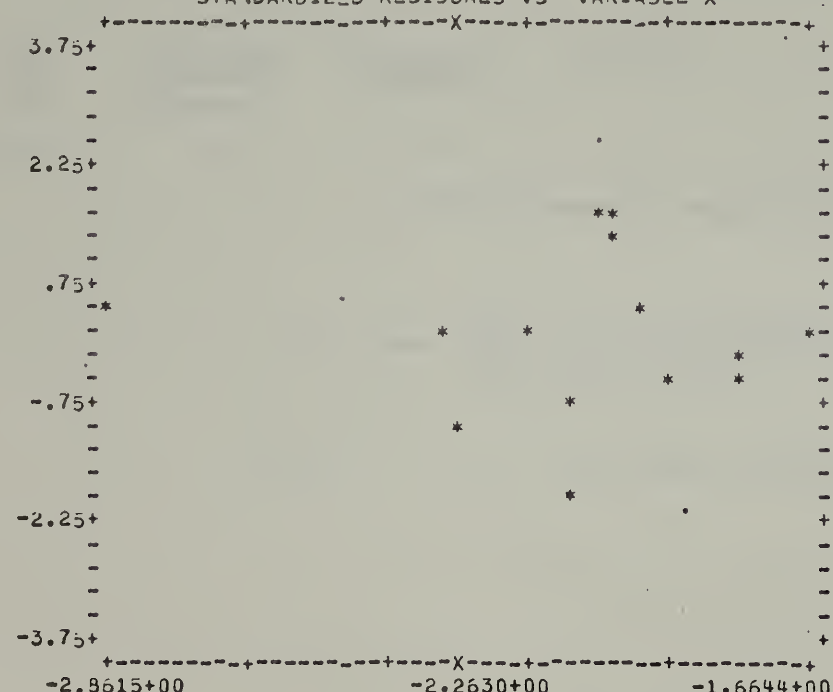
STANDARDIZED RESIDUALS VS ROW NUMBER



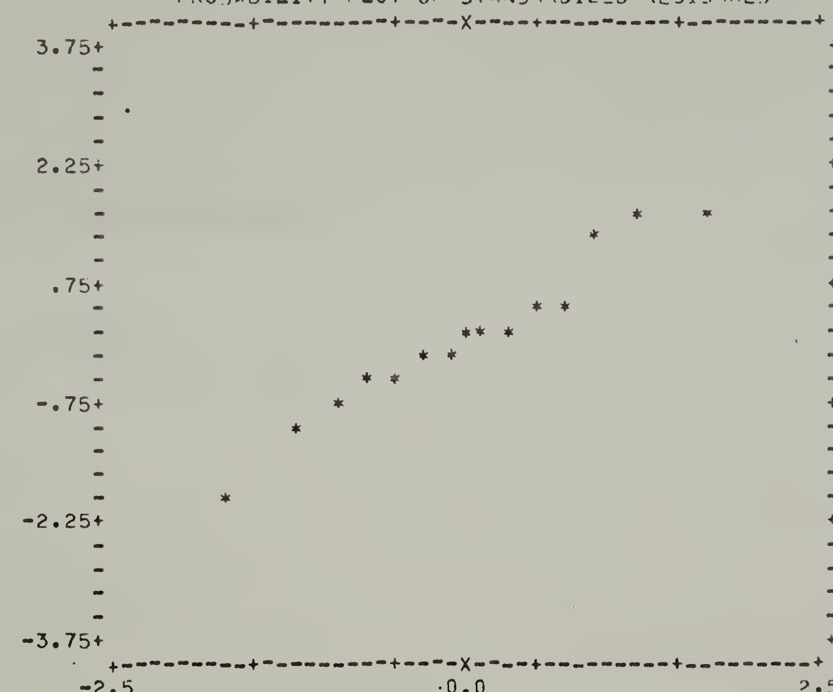
STANDARDIZED RESIDUALS VS PREDICTED VALUES



STANDARDIZED RESIDUALS VS VARIABLE X



PROBABILITY PLOT OF STANDARDIZED RESIDUALS



OMNITAB NEW HAVEN CENSUS DATA

LEAST SQUARES FIT FOR DATA IN COLUMN 32
AS A LINEAR FUNCTION OF 3 PREDICTOR VARIABLES IN COLUMNS 33, 34, 36
USING 15 NON-ZERO WEIGHTS = 1.0000000

VARIANCE-COVARIANCE MATRIX OF THE ESTIMATED COEFFICIENTS

COLUMN	33	34	36
33	.65900306		
34	.31256067	.16135980	
36	.0048561032	-.0056088332	.0061844655

ANALYSIS OF VARIANCE
-DEPENDENT ON ORDER VARIABLES ARE ENTERED, UNLESS VECTORS ARE ORTHOGONAL-

COLUMN	SS=RED. DUE TO COEF.	CUM. MEAN & REDUCTION	D.F.	CUM. RESIDUAL MS	D.F.	F(COEF=0)	P(F)	F(COEFS=0)	P(F)
33	35.819093	35.819093	1	.31260395	14	194.173	.000	58.632	.000
34	.32169908	18.070396	2	.31190432	13	1.744	.211	5.962	.017
36	1.8411130	12.560635	3	.18447027	12	9.981	.008	9.981	.008
RESIDUAL	2.2136432		12						
TOTAL	40.195548		15						

LEAST SQUARES FIT FOR DATA IN COLUMN 32
AS A LINEAR FUNCTION OF 3 PREDICTOR VARIABLES IN COLUMNS 33, 34, 36
USING 15 NON-ZERO WEIGHTS = 1.0000000

	ESTIMATES FROM LEAST SQUARES FIT				FIT OMITTING LAST COLUMN		
COLUMN	COEFFICIENT	S.D. OF COEFF.	RATIO	*ACC. DIGITS	COEFFICIENT	S.D. OF COEFF.	RATIO
33	-.29137449	.81179003	-.36	6.66	-.48645498	1.0525219	-.46
34	.29672059	.40169616	.74	7.30	.52203993	.51403125	1.02
36	.24844378	.078641372	3.16	7.65			
RESIDUAL STANDARD DEVIATION =			.42950002		.55948395		
BASED ON DEGREES OF FREEDOM			15- 3 = 12		15- 2 = 13		

* THE NUMBER OF CORRECTLY COMPUTED DIGITS IN EACH COEFFICIENT USUALLY DIFFERS BY LESS THAN 1 FROM THE NUMBER GIVEN HERE

The preceeding four pages of printed computer output are described fully in NBS Technical Note 552,

Omnitab II User's Reference Manual by D. Hogben, S. T. Peavy, and R. N. Varner, p. 152 ff, Oct. 1971.

APPENDIX B

Bibliography

1. Bachman, Gaylen D., Jacqueline M. Schwartz, and Vernon E. Weckwirth, "A Report on Lead Poisoning in Children and Youth Projects," Systems Development Project, School of Public Health, University of Minnesota, Oct. 1970.

2. Barltrop, Donald and N. J. P. Killala, "Factors Influencing Exposure of Children to Lead", Arch. Dis. Childh., Vol. 44, No. 476, Aug. 1969.

3. Bazell, Robert J., "Health Programs: Slum Children Suffer Because of Low Funding", Science, Vol. 172, May 1971.

4. Berg, J. M. and M. Zappella, "Lead Poisoning in Childhood with Particular Reference to Pica and Mental Sequelae", J. of Mental Deficiency Research, Vol. 8, June 1964.

5. Byers, R. K., "Lead Poisoning Review of the Literature and Report on 45 Cases", Pediatrics, Vol. 23, March 1959.

6. Challop, Roger, "Lead Poisoning in Children - National Evaluation Proposal", HEW, April 1971.

7. Chisolm, J. Julian, "Childhood Lead Intoxification: A Preventable Summertime Disease", Pediatrics Digest, Vol. 6, July 1964.

8. Chisolm, J. Julian, "Lead Poisoning", Scientific American, Vol. 224, No. 2, Feb. 1971.

9. Chisolm, J. Julian, "Treatment of Lead Poisoning", Modern Treatment, Vol. 4, July 1967.

10. Chisolm, J. Julian, "The Use of Chelating Agents in the Treatment of Acute and Chronic Lead Intoxification in Childhood", J. of Pediatrics, Vol. 73, No. 1, July 1968.

11. Chisolm, J. Julian and Harold E. Harrison, "The Exposure of Children to Lead", J. of Pediatrics, Vol. 18, Dec. 1956.

12. Chisolm, J. Julian and Eugene Kaplan, "Lead Poisoning in Childhood--Comprehensive Management and Prevention", J. of Pediatrics, Vol. 73, No. 6, Dec. 1968.

13. Christian, Joseph R., Bohdan S. Celewyez and Samuel L. Andelman, "A Three-year Study of Lead Poisoning in Chicago", AJPH, Vol. 54, No. 8, Aug. 1964.

14. Clark, Alan J. and George W. Hallett, "Lead Poisoning Survey--Portland Main", J. of Maine Medical Association, Vol. 62, Jan. 1971.

15. Cohen, George J. and Walter E. Ahrens, "Chronic Lead Poisoning", J. of Pediatrics, Vol. 54, No. 3, March 1954.

16. Cohen, George J., Reginald S. Lourie, and William Abernethy, "Grand Rounds: Pica and Lead Poisoning", Clinical Proceedings Children's Hospital of the District of Columbia, Vol. XXI, Nos. 7 and 8, July-Aug. 1965.

17. Conn, Robert H. and Dudley Anderson, "D.C. Mounts Unfunded Program of Screening for Lead Poisoning", HSMHA Health Reports, Vol. 86, No. 5, May 1971.

18. "Control of Lead Poisoning in Children", HEW, PHS, BCEM, Dec. 1970.

19. Deeths, Tony M. and James T. Breeden, "Poisoning in Children--A Statistical Study of 1057 Cases", J. of Pediatrics, Vol. 78, No. 2, Feb. 1971.

20. "Facts About Lead and Pediatrics", Lead Industries Association Inc.

21. Fine, Philip R., "A Lead Poisoning Surveillance Program Proposal", Division of Disease Control, Illinois Department of Public Health, May 1971.

22. Fisher, Martha, "Lead Poisoning in Children", Journal of Housing, Vol. 6, 1970.

23. Greengard, Joseph, "Lead Poisoning in Childhood: Signs, Symptoms, Current Therapy, Clinical Expressions", Clinical Pediatrics, Vol. 5, No. 5, May 1966.

24. Griggs, Robert C., Irving Sunshine, Vaun A. Newill, Burritt W. Newman, Stuart Buchanan and Cleo A. Rasch, "Environmental Factors in Childhood Lead Poisoning", J. of the AMA, Vol. 187, No. 10, March 1964.

25. Guinee, Vincent F., "Lead Poisoning in New York City", paper presented at New York Academy of Sciences, Dec. 1970.

26. Guinee, Vincent F., "The Position of New York City on the Level of Appropriations for Control of Childhood Lead Paint Poisoning", papers submitted to the Sub-committee on Labor and Public Welfare of the Senate Committee on Appropriations, July 1970 and July 1971.

27. "Hearing Before the Subcommittee on Health of the Committee on Labor and Public Welfare, U. S. Senate, Ninety-first Congress", Nov. 1970.

28. "Hearings Before the Subcommittee on Housing of the Committee on Banking and Currency", House of Representatives, Ninety-first Congress, July 1970.

29. Ingalls, Theodore H., Emil A. Tiboni, and Milton Werrin, "Lead Poisoning in Philadelphia 1955-1960", Archives of Environmental Health, Vol. 3, Nov. 1961.

30. Jacobziner, Harold, "Lead Poisoning in Childhood: Epidemiology, Manifestations, and Prevention", Clinical Pediatrics, Vol. 5, No. 5, May 1966.

31. Jacobziner, Harold and Harry W. Raybin, "The Epidemiology of Lead Poisoning in Children", Archives of Pediatrics, Feb. 1962.

32. Jacobziner, Harold and Harry W. Raybin, "Lead Poisoning in Young Children - Fatal and Nonfatal", N. Y. State J. of Med., Jan. 1960.

33. Jenkins, C. David and Robert B. Mellins, "Lead Poisoning in Children--A Study of Forty-six Cases", AMA Archives of Neurology and Psychiatry, Jan. 1957.

34. Kehoe, Robert A., "The Metabolism of Lead in Man in Health and Disease", J. of the Royal Institute of Public Health and Hygiene, 1961.

35. Kehoe, Robert A., "Normal Metabolism of Lead", Archives of Environmental Health, Vol. 8, Feb. 1964.

36. "Lead Poisoning in Children: Diagnostic Criteria", HEW, May 1959.

37. "Lead Poisoning in Children from Paint", BCEM, Aug. 1969.

38. Lin-Fu, Jane, "Childhood Lead Poisoning", Children, Vol. 17, No. 1, Jan.-Feb. 1970.

39. Lin-Fu, Jane, "Lead Poisoning in Children", HEW Children's Bureau Publication No. 452-1967, 1967.

40. Lutz, Garson A., Arthur A. Levin, Sanford G. Bloom, Kaj L. Nielsen, Jack L. Cross, and David L. Morrison, "Technical, Intelligence, and Project Information System for the Environmental Health Service, Volume III", Lead Model Case Study, Battelle Memorial Institute, June, 1970.

41. McLaughlin, Mary, "Two Health Problems--One Solution", Bull. N. Y. Acad. Med., Vol. 46, No. 6, June 1970.

42. "Medical Aspects of Childhood Lead Poisoning", HSMHA Health Reports, Vol. 86, No. 2, Feb. 1971.

43. Meigs, J. Wister and Elaine Whitmore, "Epidemiology of Lead Poisoning in New Haven Children--Operational Factors", Connecticut Medicine, Vol. 35, No. 6, June 1971.

44. Mellins, Robert B. and C. David Jenkins, "Epidemiological and Psychological Study of Lead Poisoning in Children", J. of AMA, Vol. 158, No. 1, May 1955.

45. Oberle, Mark W., "Lead Poisoning: A Preventable Childhood Disease of the Slums", Science, Vol. 165, Sep. 1969.

46. Public Law 91-695, Ninety-first Congress, H. R. 19172, Jan. 13, 1971.

47. Rait, J. R. and C. Teng, "Lead Poisoning Issue Study", New York City Rand Institute, Feb. 1970.

48. "Report of Governor's Task Force on Lead Poisoning", State of Connecticut, Dec. 1970.

49. Sachs, Henrietta K., Lorry A. Blanksma, Edward F. Murray, and Morgan J. O'Connell, "Ambulatory Treatment of Lead Poisoning, Report of 1155 Cases", Pediatrics, Vol. 46, No. 3, Sep. 1970.

50. Schucker, George W., Edward H. Vail, Elizabeth B. Kelly, and Emanuel Kaplan, "Prevention of Lead Paint Poisoning Among Baltimore Children", Public Health Reports, Vol. 80, No. 11, Nov. 1965.

51. Smith, Hugo Dunlap, "Pediatric Lead Poisoning", Archives of Environmental Health, Vol. 8, Feb. 1964.

52. "The Surgeon General's Policy Statement on Medical Aspects of Childhood Lead Poisoning", Nov. 1970.

53. "Symposium on Environmental Lead Contamination", Public Health Service Publication No. 1440, March 1966.

54. Tiboni, Emil A., "Control of Childhood Lead Poisoning", Public Health Reports, Vol. 79, No. 2, Feb. 1964.

55. Tyler, Raymond L., "Philadelphia Combats 'Silent Epidemic' in the Ghetto", J. of Environmental Health, Vol. 33, No. 1, July-Aug. 1970.

56. Williams, Huntington, Emanuel Kaplan, Charles E. Couchman and R. R. Sayers, "Lead Poisoning in Young Children", Public Health Reports, Vol. 67, No. 3, March 1952.

DEPARTMENT OF COMMERCE ★ NATIONAL BUREAU OF STANDARDS ★