Volume Title: Fatal Years: Child Mortality in Late Nineteenth-Century America

Volume Author/Editor: Samuel H. Preston and Michael R. Haines

Volume Publisher: Princeton University Press

Volume ISBN: 0-691-04268-3

Volume URL: http://www.nber.org/books/pres91-1

Conference Date: n/a

Publication Date: January 1991

Chapter Title: Distinguishing the Relative Importance of Various Social, Economic, and Residential Factors

Chapter Author: Samuel H. Preston, Michael R. Haines

Chapter URL: http://www.nber.org/chapters/c11544

Chapter pages in book: (p. 137 - 176)

FOUR

DISTINGUISHING THE RELATIVE IMPORTANCE OF VARIOUS SOCIAL, ECONOMIC, AND RESIDENTIAL FACTORS

IN CHAPTER 3, we presented information about the extent of child mortality variation among major social groupings of the American population at the turn of the century. Occasionally, we examined variation according to factors taken two or three at a time, but cell sizes often quickly became too small to produce reliable results. There was, consequently, little opportunity to consider which of the many factors reviewed are most closely associated with child mortality and which appear essentially irrelevant once other factors are taken into account. This chapter addresses these questions through multivariate analysis.

Death is a biological event, and all social influences must ultimately operate through specific biological mechanisms. Those mechanisms are not visible in the census sample, and their operations must be inferred or assumed. We are, however, able to describe relations among the socioeconomic variables contained in the data set and to see which relations appear most important and which seem to be largely by-products of differential endowments of groups with other, more important, determinants of mortality.

Analytic Strategy

The principal analytic model used in this chapter is weighted least squares multiple regression. The dependent variable in the model is the mortality index (i.e., the ratio of actual to expected child deaths) calculated for an individual woman. Each observation, however, is weighted by number of children ever born to a woman, so that women with more children in a sense contribute more observations, just as they do in the population's overall level of child mortality. Weighting by number of children ever born also reduces the problem of heteroskedasticity, the tendency for larger variance to be associated with units containing fewer observations, a tendency that violates the assumptions of ordinary least squares regression. Other research indicates that this procedure produces results that are statistically "well-behaved" and that the truncation problem at zero does not require a Tobit transformation (Trussell and Preston 1982). The independent variables are, with few exceptions, represented as sets of categories rather than as continuously valued. This representation allows us the opportunity to observe a nonlinear relation between the variable and the level of child mortality.

Initially, we pursue a sequential strategy for identifying the influence of variables on child mortality. The rationale for the sequential estimation strategy is that the values of certain variables are determined before the values of other variables, and exert an influence on those variables. If we were simply to estimate one equation that included all variables, we would have no information on these causal pathways. For example, the effect of being black on occupation, and through occupation on mortality, would not be represented; we would know only the remaining effect of being black after many of the paths through which race can affect mortality were statistically controlled. The sequential approach allows a fuller appreciation of the manifold influences on mortality, and of the relations among them.

The first equation that we estimate includes only independent variables that are inherent to a woman and cannot be affected by her behavior—her race, nativity, and age. These are background variables determined at the woman's birth. The sign of the age variable is expected to be positive. Older mothers experienced (on average) a higher child mortality level since, as shown in Chapters 2 and 3, child mortality was declining in the United States in the last two decades of the nineteenth century. Thus, mother's age acts as a partial control for the time trend in mortality.

The second equation adds variables representing a woman's literacy (in any language) and her ability to speak English. These characteristics are typically acquired in childhood or adolescence, before marriage and the birth of the woman's children. They reflect primarily conditions in the family in which she was raised. The third equation then adds similar basic characteristics of the husband that are "acquired" by the woman at marriage—his literacy and ability to speak English.[1] The expected signs of these variables were discussed in Chapter 3. We estimated a version of the third equation that included the husband's nativity, but this variable was too highly collinear with wife's nativity to allow robust estimates of the joint effects of wife's and husband's nativity.

The final equation then adds all remaining characteristics that were not permanent to either husband or wife: region of current residence, size of place of residence, property ownership, wife's labor-force activity, migration status, husband's employment, husband's age and occupation (or occupational income), and presence of servants or boarders in the household. The expected signs of most of these variables have already been discussed. Husband's age is treated differently than mother's age because it is believed to serve in part as a proxy for his income, rather than for trends in child mortality.

Some of the variables described in Chapter 3 are modified or augmented to take advantage of the greater statistical power afforded by multivariate analysis. Instead of the five census regions used in most of the tabular presentations, the regression analysis uses the nine census subregions. The property-ownership variable, whether a farm or home was owned or rented, has been supplemented with information about whether the farm or home was owned free and clear or mortgaged. Outright ownership may imply higher personal net worth, which should be reflected in lower child mortality. Wife's labor-force status was altered to add information on the availability of child care in the household: in particular, whether a servant or adult female relative (aged 16 or over) was present in the household. It is expected that availability of child care will reduce child mortality, especially for working women. The combination of information on whether a woman is working and whether she has child care available in the home produces a variable with four categories.

Two new migration-status variables have been constructed, representing the proportion of a woman's life that was lived in the United States before age 14 and the proportion lived in the U.S. after age 14. These variables are constructed from information on a woman's year of immigration to the U.S. It is expected that longer residence in the U.S. would increase social integration and reduce child mortality, since most of the immigrants came from higher mortality areas in Europe. The integration process is expected to be fastest for those arriving before age 14, so a larger coefficient is expected on this variable.

As mentioned above, the 1900 census contained no information on income, a very unfortunate omission that was not corrected until the 1940 census. We have gone to considerable lengths to fill in this gap. In particular, we have constructed two variables based upon the husband's occupation. One is the mean income of the husband's (detailed) occupation, and the other is the mean number of months unemployed during the period June 1, 1899 to May 31, 1900, for men in that occupation. Construction of these variables is described in Appendix A. We did not attempt any income assignment for farmers, since farmers' incomes depended upon weather, location, soil, farm size, and capital available, so that any income that we could assign to farmers would contain enormous measurement error.

We have also constructed an index of state income, or, more precisely, earnings levels. Construction of this variable is described in Appendix B. It is designed to be an index that is independent of the occupational composition of a state, since we already have information on the husband's occupation. We use this index in both price-adjusted and price-unadjusted form. Logarithmic transformations of income variables are used because we expect that income will show diminishing returns in its impact on child mortality, as it does in international comparisons at the turn of the century and later (Preston 1975). We experimented with polynomials to represent income, but results were not appreciably different from those where logarithms were used. Husband's age and husband's age squared were included to account for the observed curvilinear (parabolic) shape of male age/earnings profiles in this period (Appendix A).

The basic data set used in this chapter is smaller than that used in most of Chapter 3. Because we are interested in the impact of many of the husband's characteristics on child mortality, we have eliminated observations in which the husband was absent from the household. We have also deleted cases for which information was missing on key variables. Finally, we introduce the restrictions used in the analysis of trends in Chapter 3 to ensure that marital duration reflects as accurately as possible children's exposure to the risk of death. The final data set includes 10,369 currently married women with husbands present, compared to the set of 13,429 women used in most of Chapter 3.

Within a set of categories such as occupation, the selection of a particular category to be the "reference category," against which the child mortality impact of other categories is measured, is fundamentally arbitrary. It does not affect either the amount of variance explained by the set or the quantitative differences between coefficients within the set. Largely for expositional purposes, we have chosen large groups, often with extreme values of child mortality, to be the reference categories.

Results for All Groups Combined

The basic results of this chapter are presented in Table 4.1. The equations in Table 4.1 are all jointly significant at a 1 percent level, as indicated by the F-ratios. The adjusted R^2 values are not large, the highest being .051 in equation (4) in Table 4.1. It is not unusual to find this level of explanatory power in models fit to individual-level data, however, since much individual variation is idiosyncratic and unaccounted for by quantifiable variables. Grouped data yield much higher adjusted R^2 values using some of the same explanatory variables (see Chapter 5), since "random" variation tends to average out in a large group.

Race

Higher child mortality for black women is retained in multivariate analysis, as shown by the large positive and statistically significant coefficients that appear in Table 4.1. Some of the effect of being black operates through other socioeconomic variables, since the coefficient for black women is reduced when successive waves of independent variables are introduced. In equation (1) of Table 4.1, where only nativity and a woman's age are controlled, black women have 51 percent higher child mortality than white women (i.e., the coefficient for black women is .506, and highly significant). Introducing literacy and English-speaking ability of the woman in equation (2) reduces the effect of race alone by 15 percent, to .433. Additional controls for literacy, English ability, and nativity of husband in equation (3) have only a small effect, reducing the coefficient by only 6 percent. Finally, when variables representing region, size of place, property ownership, mother's labor force status, migration status, husband's unemployment and occupation, and presence of boarders and servants are introduced in equation (4), the race coefficient declines to .297. Thus, about 41 percent of the effect of race on child mortality can be accounted for by its impact on other measured variables that are also associated with mortality.

In theory, controlling for all relevant socioeconomic and demographic factors should eliminate entirely the effect of race, since there is no reason to believe that genetic differences were responsible for significantly greater biological frailty among black children, despite views of the time. But, in the present case, the selection of variables was limited and thorough controls for such factors as quality of housing, environment, water supply, sewerage disposal, clothing quality, and nutrition and food purity were not available. Moreover, many of the variables that we are including undoubtedly had a different meaning for blacks and whites. Blacks in cities were residentally segregated and fell outside the reach of most ameliorative social programs (Katz 1986). Southern rural blacks lacked the property re-

TABLE 4.1

Equations Predicting the Ratio of Actual to Expected Child Deaths for Individual Women: U.S., 1900

Independent variables	(1)		(2)		(3)		(4)	
	Coefficient	*Significance*	*Coefficient*	*Significance*	*Coefficient*	*Significance*	*Coefficient*	*Significance*
Intercept	0.5840	***	0.5899	***	0.5880	***	0.6280	***
Race of wife								
White	NI	NI	NI	NI	NI	NI	NI	NI
Black	0.5060	***	0.4327	***	0.4077	***	0.2968	***
Nativity of wife								
Native with native mother	NI	NI	NI	NI	NI	NI	NI	NI
Native with British mother	−0.0229	—	−0.0136	—	−0.0081	—	−0.0345	—
Native with Irish mother	0.1737	***	0.1853	***	0.1899	***	0.0562	—
Native with Scandinavian mother	−0.1981	—	−0.1869	—	−0.1809	—	0.0664	—
Native with German mother	0.0030	—	0.0142	—	0.0194	—	−0.0490	—
Native with other West European mother	−0.4059	***	−0.3888	**	−0.3903	**	−0.3055	*
Native with East European mother	−0.1855	—	−0.2020	—	−0.2034	—	−0.1479	—
Native with South European mother	−0.4914	—	−0.4782	—	−0.4723	—	−0.4353	—
British	0.2080	**	0.2120	**	0.2168	**	0.0592	—
Irish	0.3648	***	0.3678	***	0.3694	***	0.1169	—
Scandinavian	0.0810	—	0.0742	—	0.0876	—	0.1290	
German	0.2236	***	0.2114	***	0.2162	***	0.0923	—
Other West European	−0.0838	—	−0.1059	—	−0.1264	—	−0.2089	—
East European	0.1869	***	0.0950	—	0.0860	—	−0.0779	—
South European	0.3120	***	0.1702	—	0.1482	—	−0.0816	—
Other foreign-born	0.2297	***	0.1980	***	0.1910	***	0.0443	—
Age of wife	0.0069	***	0.0063	***	0.0062	***	0.0094	***

Literacy of wife						
Literate	NI	NI	NI	NI	NI	NI
Illiterate	0.1520	***	0.1065	**	0.1004	**
Wife speaks English						
Yes	NI	NI	NI	NI	NI	NI
No	0.1236	*	0.0282	—	0.0029	—
Literacy of husband						
Yes			NI	NI	NI	NI
No			0.1190	***	0.1234	***
Husband speaks English						
Yes			NI	NI	NI	NI
No			0.2090	**	0.1506	—
Region						
New England					NI	NI
Middle Atlantic					−0.0557	—
East North Central					−0.0928	—
West North Central					−0.1302	**
South Atlantic					−0.2235	***
East South Central					−0.1641	*
West South Central					−0.0074	—
Mountain					0.4308	***
Pacific					−0.0615	—
Residence						
Top 10 cities					0.2954	***
Other cities 25,000+					0.3160	***
Cities 5,000–24,999					0.2040	***
Cities 1,000–4,999					NI	NI
Rural					0.0682	—

TABLE 4.1 *(cont.)*

Independent variables	*(1)*		*(2)*		*(3)*		*(4)*	
	Coefficient	*Significance*	*Coefficient*	*Significance*	*Coefficient*	*Significance*	*Coefficient*	*Significance*
Property								
Owns farm clear							−0.1577	**
Owns farm mortgaged							−0.1005	—
Rents farm							−0.0634	—
Owns home clear							−0.0417	—
Owns home mortgaged							−0.1064	**
Rents home							NI	NI
Wife working								
Working—has child care							0.2139	—
Working—no child care							0.0357	—
Not working—has child care							0.0139	—
Not working—no child care							NI	NI
Migration status								
Proportion of life lived in the U.S. before age 14							−0.0266	—
Proportion of life lived in the U.S. after age 14							0.0186	—
Other dichotomous variables								
Husband unemployed							0.1586	***
Servant(s) in household							−0.0211	—
Boarder(s) in household							0.1510	***
Income								
Husband's age							−0.0006	—
Husband's age squared							0.0000	—
Log of state income index							−0.4854	***

Husband's occupation								
Professional, Technical							−0.0652	—
Agricultural (excluding Laborers)							−0.1454	**
Agricultural Laborers							−0.0465	—
Managers, Official, Proprietors							−0.1098	*
Clerical							−0.1650	*
Sales							−0.2093	**
Craftsmen, Foremen							−0.0217	—
Operatives							−0.0904	*
Service Workers							−0.1189	—
Laborers							NI	NI
Miscellaneous							−0.1424	—
N	9,917		9,917		9,917		9,917	
Adjusted R-squared	0.0238		0.0257		0.0267		0.0505	
F-ratio	13.727	***	12.387	***	11.071	***	8.875	***

Source: Sample of census enumerators's manuscripts, U.S. 1900.

Note: Sample consists of currently married women with husband present, married less than 25 years. The dependent variable is the ratio of actual to expected child deaths for each woman. For an explanation of its construction, and of the other variables, see text. Regressions are weighted least squares with the weights being the number of children ever born (normalized back to the original number of observations).

Significance levels are: *** = significant at least at a 1 percent level; ** = significant at least at a 5 percent level; * = significant at least at a 10 percent level; — = not significant at least at a 10 percent level. NI = not included.

sources of southern rural whites, and any literacy they attained came through poorer public schools. Even though they may have overlapped with whites in the occupational distribution, blacks were typically paid less within a particular occupation (Wright 1986; Ransom and Sutch 1977). That husband's income may be a key omitted variable is suggested by Woodbury's (1925) results for eight cities during 1911–15. When income differences were controlled by standardization, the black disadvantage in infant mortality was only 9 percent (Woodbury 1925:122).

Nativity of Wife

The higher mortality of foreign-born women largely disappears in multivariate analysis. Signs of the coefficients among this group of dummy variables denote mortality levels relative to that in the reference category, native-born women with native-born mothers. In equation (1) of Table 4.1, seven of the eight categories of foreign-born women had child mortality that was higher than that of native whites of native parentage, and most differences were statistically significant. In addition, native-born women with Irish mothers also had significantly higher child mortality than native-born women of native parentage. For second-generation mothers in general, however, mortality was not significantly different from that of native born women of native parentage, and in fact it was typically somewhat lower.

As additional independent variables are added in equations (2)–(4), the importance of nativity diminishes, indicating that the effect of nativity, unlike that of race, largely operated through other measured variables. In equation (1) of Table 4.1, eight out of fifteen nativity variables are significant at least at a 5 percent level; but none of the fifteen remains significant in equation (4). Many of the coefficients also are reduced in magnitude, especially when they were large in equation (1) (e.g., women born in Ireland, Germany, or Britain). Almost none of the decline in coefficients occurs until equation (4), implying that the child mortality disadvantage of foreign-born women was not a result of literacy or English-speaking disabilities among them or their husbands, but rather of more immediate economic and residential characteristics.

The second-generation immigrants did not suffer from abnormally high mortality even before controls for current economic circumstances are instituted. When such controls are introduced, the second-generation mothers typically have child mortality levels below those of native white women of native parentage (i.e., their coefficients are generally negative in equation [4], although none of the differences is significant). In other words, whatever ethnic differences in child-care practices and breastfeeding were present among first- and second-generation mothers, they were not important enough to have left a vivid mark on child mortality levels by ethnicity, once other measured factors are controlled. In particular, breastfeeding differences that were noted at the time and that we described in Chapters 1 and 3 do not appear to have left a deep imprint on child mortality. Foreign-born mothers did have higher child mortality, but the principal causes are to be found in their economic and social circumstances rather than in unmeasured variables that ethnicity may reflect.

Probably the most important of these economic and social variables in accounting for the 12 percent higher child mortality of foreign-born mothers shown in Table 3.1 is urban residence. The distribution of the married women used to calculate the child mortality index is presented in Table 4.2. Foreign-born white women were overwhelmingly (71.8 percent) concentrated in urban areas, and, of those living in cities, 72 percent were living in the unhealthiest cities of 25,000 or more inhabitants. Indeed, 42 percent were resident in the 10 largest urban centers. In contrast, among native white women with native mothers, 64.8 percent were living in rural areas, and of those who did reside in cities, over half (58.2 percent) were living in the smaller, healthier urban areas below 25,000 in population. Child mortality of native white women of foreign parentage was intermediate between the native women of native parentage and foreign-born white women, but their residence pattern was much more similar to that of the foreign born. About two-thirds (62.2 percent) were in urban areas, and, of those, 64 percent lived in the cities of over 25,000 inhabitants. Thus, the residence patterns of first- and second-generation immigrant families placed them in a disadvantageous position vis-à-vis child survival. But once residential and socioeconomic variables are controlled, their disadvantage is essentially eliminated.

Age of Wife

The age of the wife was introduced to control for the downward trend in mortality. Older mothers bore children at earlier dates, on average, and their children would have been exposed to higher mortality conditions. The uniformly positive and significant coefficients on this variable in Tables 4.1 are consistent with the existence of such a time trend and justify the inclusion of this variable as a correction for the trend. The value of this variable of .01 in equation (4) suggests that child mortality levels were raised by about 1 percent for each

TABLE 4.2
Women and Children Ever Born in the Sample Used to Calculate the Child Mortality Index, by Race, Nativity, and Residence

	Urban	*Top 10 cities*	*Other cities 25,000+*	*Cities 5,000–24,999*	*Cities 1,000–4,999*	*Rural*	*Unknown*	*Total*
Women								
Total (*N*)	4,670	1,322	1,317	1,039	992	5,614	85	10,369
White	4,439	1,297	1,240	979	923	4,809	69	9,317
Native white	2,950	668	794	727	761	4,238	56	7,244
Native mother	1,924	292	513	519	600	3,623	47	5,594
Foreign mother	1,026	376	281	208	161	615	9	1,650
Foreign white	1,489	629	446	252	162	571	13	2,073
Black	224	21	75	60	68	770	15	1,009
Percentage of total								
Total	45.0	12.7	12.7	10.0	9.6	54.1	0.8	100.0
White	47.6	13.9	13.3	10.5	9.9	51.6	0.7	100.0
Native white	40.7	9.2	11.0	10.0	10.5	58.5	0.8	100.0
Native mother	34.4	5.2	9.2	9.3	10.7	64.8	0.8	100.0
Foreign mother	62.2	22.8	17.0	12.6	9.8	37.3	0.5	100.0
Foreign white	71.8	30.3	21.5	12.2	7.8	27.5	0.6	100.0
Black	22.2	2.1	7.4	5.9	6.7	76.3	1.5	100.0
Percentage of urban population								
Total	100.0	28.3	28.2	22.2	21.2	—	—	—
White	100.0	29.2	27.9	22.1	20.8	—	—	—
Native white	100.0	22.6	26.9	24.6	25.8	—	—	—
Native mother	100.0	15.2	26.7	27.0	31.2	—	—	—
Foreign mother	100.0	36.6	27.4	20.3	15.7	—	—	—
Foreign white	100.0	42.2	30.0	16.9	10.9	—	—	—
Black	100.0	9.4	33.5	26.8	30.4	—	—	—
Children ever born								
Total (*N*)	15,872	4,630	4,423	3,419	3,400	21,789	314	37,975
White	15,023	4,562	4,175	3,160	3,126	18,199	251	33,473
Native white	8,992	2,066	2,392	2,078	2,456	15,608	192	24,792
Native mother	5,543	809	1,427	1,392	1,915	13,325	161	19,029
Foreign mother	3,449	1,257	965	686	541	2,283	31	5,763
Foreign white	6,031	2,496	1,783	1,082	670	2,591	59	8,681
Black	831	60	243	259	269	3,452	62	4,345
Percentage of total								
Total	41.8	12.2	11.6	9.0	9.0	57.4	0.8	100.0
White	44.9	13.6	12.5	9.4	9.3	54.4	0.7	100.0
Native white	36.3	8.3	9.6	8.4	9.9	63.0	0.8	100.0
Native mother	29.1	4.3	7.5	7.3	10.1	70.0	0.8	100.0
Foreign mother	59.8	21.8	16.7	11.9	9.4	39.6	0.5	100.0
Foreign white	69.5	28.8	20.5	12.5	7.7	29.8	0.7	100.0
Black	19.1	1.4	5.6	6.0	6.2	79.4	1.4	100.0
Percentage of urban population								
Total	100.0	29.2	27.9	21.5	21.4	—	—	—
White	100.0	30.4	27.8	21.0	20.8	—	—	—

TABLE 4.2 (*cont.*)

	Urban	*Top 10 cities*	*Other cities 25,000+*	*Cities 5,000–24,999*	*Cities 1,000–4,999*	*Rural*	*Unknown*	*Total*
Native white	100.0	23.0	26.6	23.1	27.3	—	—	—
Native mother	100.0	14.6	25.7	25.1	34.5	—	—	—
Foreign mother	100.0	36.4	28.0	19.9	15.7	—	—	—
Foreign white	100.0	41.4	29.6	17.9	11.1	—	—	—
Black	100.0	7.2	29.2	31.2	32.4	—	—	—

Source: Sample of census enumerators' manuscripts, U.S., 1900. Numbers are slightly larger than those in Table 4.1, which deletes some cases with unknown values.

additional year of mother's age, once other attributes are controlled. It is likely that the age control has reduced the impact of being foreign-born, since foreign-born women were, in general, older than natives in the sample. Indeed, the nativity differences in equation (1) of Table 4.1, where only age is controlled, are generally lower than those of Chapter 3 (though many remain significant until other variables are introduced).

Literacy and English Ability of the Wife

Literacy and ability to speak English were shown to be associated with lower child mortality in Chapter 3. Multivariate analysis casts a good deal of light on the paths by which these variables operate. When it is first introduced into the regression equation at stage 2, a woman's English ability has a relatively strong effect on child mortality, with non-English speakers having about 12 percent higher mortality. Virtually all of this effect seems to operate through the type of husband a woman marries. When husband's literacy and English-speaking ability are introduced at stage 3, the woman's own English-speaking ability loses most of its ability to differentiate child mortality levels in Table 4.1. That is, a wife unable to speak English was more likely to have an illiterate and/or non-English speaking husband. Patterns of assortative mating appear to account for the initial impact of a woman's English ability on child mortality, and no significant "direct" effect remains after the characteristics of her husband that were evident at marriage are controlled.

The mother's literacy, on the other hand, does not show such a drastic change in the size of the coefficient, and its coefficient remains statistically significant, declining by about one-third as other variables are introduced. Although literacy is an imperfect measure of education, it remains a significant predictor of child mortality. The remaining differential of 10 percent between literate and illiterate mothers is not large, however, and, as we argued in Chapter 3, some of it is likely attributable to selectivity factors; illiterate women in 1900 were probably a group that had, on average, a high incidence of other debilities. Given the relatively primitive state of health knowledge and technology in the United States in 1900, it is perhaps not surprising that the payoff to literacy was relatively low. We show in Chapter 5 that the effect of literacy is considerably greater in the recent experience of developing countries.

Literacy and English Ability of the Husband

Equations (3) and (4) of Table 4.1 add the husband's literacy and capacity to speak English to the regression model. The signs of the coefficients of these variables are positive (that is, illiteracy and inability to speak English raise child mortality). The coefficient of husband's literacy remains significant and changes little in size when other socioeconomic variables are added in the fourth equation. A husband's literacy is more important than that of his wife, a finding that we subsequently show to be primarily traceable to urban areas.

The ability of the husband to speak English is not significantly related to child mortality, although the coefficient is sizable in Table 4.1. The English ability coefficient for the husband diminished considerably between equations (3) and (4), indicating that the effect of this variable largely expressed itself through the husband's present socioeconomic and residential status.

The final equation (4) of Table 4.1 introduces the last set of socioeconomic variables, those representing the current socioeconomic and residential situation of husbands and wives.

Size of Place and Region of Residence

The very large mortality differences by size of place that appeared in Chapter 3 are retained in multivariate analysis. We have chosen the reference category to be places of 1,000–4,999 inhabitants. These places were insignificantly different from rural areas in their effects on child mortality, once other variables were controlled. Relative to these very small cities, however, cities of 5,000–24,999 had mortality levels that were elevated by 20 percent, and larger cities had about 30 percent higher mortality. As in Chapter 3, there is a slight tendency for the 10 largest cities to have lower mortality than the next largest category. Unlike many of the variables studied in this chapter, each of the categories of this variable contained a substantial proportion of the population (see Table 4.2), so that size of place had both a large and a pervasive influence on mortality levels at the turn of the century. In our view, this influence reflects primarily the greater efficiency with which infectious diseases are spread in denser areas. This efficiency operates both through direct person-to-person contact and through contamination of common resources such as the water and food supply. The twentieth century witnessed major assaults on both of these modes of transmission, but the battle was just beginning, and the principal weapons for it being prepared, in the 1890s.

Regional differences in mortality are products of many factors: differences in population composition with respect to individual characteristics such as race and literacy; differences in the macroeconomic environment that affect levels of wages and income; ecological differences in climate and geography that affect the prevalence of infectious disease; and differences in the availability of health care and preventative public-health measures. As we have argued in Chapter 1, this latter factor is probably not particularly salient in the 1890s except in a few localities, because the state of medicine and public health was simply too primitive to have sharply differentiated among the mortality levels of different places. For example, it is not clear that births supervised by physicians had better outcomes than those supervised by midwives. Hospitals were clearly to be avoided except possibly in extreme emergencies.

Recall from Chapter 3 that New England and the Mountain region had relatively high mortality and the Midwest (East North Central and West North Central) had low mortality. The South was a region of unexpectedly low mortality, largely, it appeared, because of its rural character. It is now possible to see how this regional pattern is modified when particular features of these areas and their residents are controlled.

Column 1 of Table 4.3 reproduces the original interregional differences from Chapter 3 when no other variables are controlled. Column 2 presents these differences when all of the individual-level variables in equation (4) of Table 4.1 are controlled, including an individual's size of place of residence. A very different regional pattern emerges. The Midwest remains a low mortality region, but New England is no longer the region of highest mortality. Its mortality is exceeded by that of four other regions, including the three regions of the South and Southwest. New England's debilities appear to have arisen largely from the fact that it was the most highly urbanized region. Once rural/urban residence is controlled (along with other variables, such as literacy, on which New England was *not* particularly disadvantaged), its mortality does not appear exceptionally high. It

TABLE 4.3
Regional Pattern of Child Mortality with and without Control for Various Personal and Regional Characteristics: U.S., 1900

		Coefficients of regional dummy variables[a]		
Region	*Original difference*[b]	*No control for state economic conditions*	*State income controlled*	*State income and price level controlled*
New England	NI	NI	NI	NI
Middle Atlantic	−.0868	−.0247	−.0557	−.0375
East North Central	−.2372	−.0508	−.0928	−.0512
West North Central	−.3395	−.0881	−.1302[d]	−.0835
South Atlantic	−.0871	.0117	−.2235[e]	−.1544[e]
East South Central	−.1078	.0504	−.1641[c]	−.0774
West South Central	−.0002	.1465[d]	−.0074	.0382
Mountain	−.0416	.2534[d]	.4308[c]	.3553[c]
Pacific	−.3151	−.1786[e]	−.0615	−.1221

Source: Sample of census enumerators' manuscripts, U.S., 1900.
[a] The regional dummy variables relate a region's child mortality to that of New England, controlling individual-level characteristics.
[b] The original difference in the child mortality index between the region in question and New England.
[c] Significant at a 1 percent level.
[d] Significant at a 5 percent level.
[e] Significant at a 10 percent level.

is replaced by the Mountain region, which was highly rural but which nevertheless maintained high mortality conditions.

Some indication of the conditions that may have given rise to the Mountain region's high mortality are provided by Children's Bureau study of child health in Montana in 1917 (Paradise 1919). The study describes circumstances of extreme isolation, great distances to towns and neighbors over roads that were often impassable, combined with great congestion within the largely one-room cabins. Even so, child mortality levels, though high for rural areas, were below those of each of the eight cities investigated by the Bureau. When we examine the mortality levels of specific states (not shown), those with the highest levels in the Mountain region are Arizona, Colorado, and especially New Mexico. Arizona and New Mexico had a high percentage of American Indians, and the high mortality of the region is probably traceable primarily to this group. (For a study of the very high child mortality among American Indians in the 1900 census, see Johansson and Preston 1979).

The regional pattern of mortality differences is altered once again when we control for the economic circumstances of each state. As shown in Appendix B, each of the New England states had above-average earnings levels, while the South had below-average levels (see also Wright 1986). Table 4.1 shows that (the log of) a state's income (i.e., earnings level) had a highly significant, negative impact on the child mortality level of residents. Column 3 of Table 4.3 presents the regional differences that remain after state income levels are controlled. New England again becomes a high mortality region, surpassed only by the Mountain region, an area with even higher incomes. The South once again becomes a low mortality region, significantly so in the case of the South Atlantic. In other words, New England appears to have had better mortality levels than the South once individual characteristics are controlled because it was an area of above-average income levels. Its advantage over the South, and that of the Middle Atlantic states as well, was principally economic.

A state's income level in column 3 was not adjusted for possible differences in the cost of living among states. Haines (1989a) has produced a cost of living index for states in 1890, and our final step is to see how using real income (1900 state income divided by the 1890 cost-of-living index) modifies the regional picture. Column 4 of Table 4.3 shows that New England's relatively high cost of living has overstated somewhat its economic advantage and, hence, its mortality disadvantage. It still has higher mortality than the Midwest or South, net of all other factors, but the excess is diminished. The lowest mortality region continues to be the South Atlantic, and the highest, the Mountain region. There is a 51 percent difference in child mortality levels between these regions. For all other regions the range of child mortality levels is only a modest 16 percent.[2]

To summarize, the raw differences among the child mortality levels of different regions reflect a complex array of factors. New England had the advantage of relatively high income and a low proportion of black residents, and the disadvantage of being highly urban. The South was the mirror image of New England in these respects. The outcome was a level of child mortality in the South (South Atlantic and East South Central) that was 9–11 percent lower than that of New England. When urbanization and percentage of black residents are controlled, the South had higher mortality than New England by

1–5 percent; when state income levels are additionally controlled, the South returns to having lower mortality by 8–22 percent. So when all of these characteristics are factored in, the South appears to be relatively healthy. We are surprised by this result since we expected the ecological and climatic features of the South, the effects of which should become increasingly manifest as socioeconomic characteristics are controlled, to produce above-average mortality. But it is possible that the lower population density of the South in rural areas, a factor not accounted for in our analysis, has overcome any disadvantages of climate. We return to this issue below.

The unusually favorable child mortality of the Midwest appears to be largely explicable in terms of social and economic conditions of its residents. Child mortality was 15–34 percent lower in the Midwest than in New England and the Middle Atlantic states, but after the social and economic characteristics of residents are controlled its advantage is only 3–9 percent. Undoubtedly, the rural character of the region and the high percentage of farmers were two of the main factors accounting for its lower mortality. Aggregate income differences were not significant in this regard, since the Midwest's position relative to the Northwest was largely unaffected by whether state income is controlled.

A final word is in order on the effect of state income. Since data on an individual's occupation were available in the data set, we sought a state income index that was independent of the occupational distribution in order to characterize the aggregate labor market in a state. The index pertains to 1899–1900 and is correlated with Easterlin's (1957) index of state per capita income in 1899–1900 at .810. This variable was significant in the regression equations and properly signed: a higher average annual earnings level in a state led to lower child mortality. The coefficient of −.49 in Table 4.1 implies that a 10 percent increase in annual earnings would reduce child mortality by 4.9 percent. Since annual earnings varied by a factor of 2.1 between Georgia, the lowest wage state, and Massachusetts, one of the highest, it is clear that regional variation in economic circumstances accounted for a good deal of variation in child mortality. The higher-than-average wage levels in the Northeast offset some of the disadvantage that living in this region otherwise fostered. The coefficient of state income adjusted for price levels is −.38 and is also highly significant.

Husband's Occupation and Occupational Income

For the main analysis of mortality variation by husband's occupations, we used the aggregate set of eleven occupation categories based upon the U.S. Census Bureau's 1950 classification of occupations. The reference category chosen was that of laborers, a large, low-status, high-mortality group. Results in Table 4.1 show that other occupational groups typically had lower child mortality levels than laborers. But interoccupational differences tended to be quite small. The largest disparity occurred between laborers and sales workers, but the difference in child mortality between these groups was only 20 percent. For a variable that is probably our single best indicator of socioeconomic status, this range seems very narrow. Within it, furthermore, professional and technical workers had relatively high mortality, only 7 percent below that of laborers. Farmers continued to have below-average mortality, even though their predominantly rural residence was controlled. Only farmers and sales workers had mortality that differed from that of laborers at a 5 percent significance level.

One might suspect that the occupational categories used in Table 4.1 are too coarse to capture the significant variation that existed, despite the fact that even coarser categories used in England and Wales around the same time show much larger variations in child mortality (see Chapter 5). In part to address this concern, and in part to tie occupation more closely to economic standing, we created a variable representing the predicted income of an occupational group. Incomes were assigned based on the detailed, three-digit occupational code from the 1900 census rather than the cruder categories used above. Details of the index construction are presented in Appendix A. Note that incomes of professionals, managers, and officials were generally quite a bit higher than those of laborers and other blue-collar workers, so that the negligible mortality differences between the groups just described are not attributable to anomalous economic disparities of the time (though income differentials do appear to be smaller than at present). Salesmen and merchants did well economically, on average, so that their favorable child mortality might have an economic basis. On the other hand, domestic-service employees and those in agricultural pursuits had quite low incomes, in general, which were not reflected in unusually high child mortality.

From these scattered observations, it is already clear that our predicted income variable is not likely to perform well. Nor does it. When we use this variable in place of husband's occupation in a subsample excluding farmers (for whom no income estimate was attempted), its coefficient is .023, only about 5 percent of the coefficient on state income, and it has the wrong sign. Women whose husbands had higher-income occupations had very slightly, and insignificantly, increased child mortality.

It is not likely that husband's income was unimportant for mortality. Woodbury's (1925) study for the Children's Bureau demonstrated strong income effects on infant morality in cities, and our own index of state earnings levels is strongly related to child mortality. Evidently, there is too much measurement error in the variable for it to be useful. One source of the error may be in the occupational income levels themselves, which are culled from different sources. We created an interactive variable (log of predicted income times a dummy variable that took the value of one if the source was the 1901 Cost of Living Survey and zero otherwise) to see whether source differences were important. But the coefficient of this variable is only .006 and insignificant. Another source of measurement error, probably more important, is simply that there was substantial income variance within occupations, variance not accounted for when all members of that occupational group are assigned the same mean income. With enough variance, the average income of an occupation becomes essentially meaningless. This appears to be the case.

Husband's Unemployment and Unemployment Levels in His Occupation

Unemployment in 1900 was, as now, a sensitive indicator of economic distress in the family. Indeed, it was more serious at the turn of the century, since there was little in the way of a "safety net"; unemployment insurance, welfare benefits, and private charity were meager if they existed at all (Patterson 1986: ch. 2). Table 3.1 indicated that almost 18 percent of the households with both spouses present (and for whom information was reported) experienced some unemployment in the twelve months before the census of June 1, 1900. The census year was not a year of relative economic difficulty or particularly high unemployment. Unemployed workers as a percentage of the civilian labor force were only 6.5 percent in 1899 and 5.0 percent in 1900. This level was well below the average of 14.2 percent that had prevailed during the period 1893–98 (U.S. Bureau of the Census 1975: Series D 85–86). It must be remembered that bouts of unemployment were a fact of life for many manual workers, especially those in the building trades.

A strong relation between child mortality and husband's unemployment in the census year was observed in Table 3.1. The coefficient of this variable is also large and statistically significant in Table 4.1. Unemployment of the principal income-earner in the household at some time during the year before the census was associated with a 16 percent increase in child mortality. In all likelihood, this coefficient is so large not because of the impact of unemployment in 1899–1900 per se but because unemployment in the census year is statistically associated with unemployment in previous years as well. Recall that impoverished families in Washington, D.C., in 1901 complained not so much about the level of wages that they were paid but about the irregular availability of work.

To examine whether unemployment levels in the occupation currently pursued by the husband contributed to explaining variation in mortality, over and above the contribution of the husband's own unemployment, we created an additional variable that is described in Appendix A. This variable is the average number of months unemployment for men in the husband's three-digit (i.e., detailed) occupational category. When this variable is added to equation (4) in Table 4.1 (not shown), its coefficient is −.0074. It is insignificant and, like occupational income, takes the wrong sign. The husband's own recent history of unemployment proves to be far more important than the level of unemployment in his occupation. Hence, our considerable efforts to use information about the husband's occupation to characterize the economic circumstances of the household have proven largely fruitless.

Wife's Labor-Force Participation

Earlier tabulations revealed a complex relationship between child mortality and the labor-force activity of the mother. The one-way comparison (Table 3.1) showed considerably higher child mortality among working mothers, but the relation did not pertain to native white women, and was reversed among households with husbands absent. The regressions in this chapter exclude women with absent husbands, and the negative "effect" of mother's work on child mortality is maintained. In Table 4.1, wife's labor-force activity was combined with a variable pertaining to child-care availability: whether there was another nonboarding female adult in the household who could potentially have provided child care. For women with child care in the home, child mortality was 20 percent higher among working than among nonworking women. But for women without child care, where the effect of working might be expected to be greatest, the differential was only 3.6 percent. None of the differences are statistically significant. Below, we show that the anomalous result involving child-care availability was confined to rural areas and that the anticipated result did pertain in urban areas.

Property Ownership

The basic relationships involving ownership variables that were observed in Tables 3.1 and 3.5 are also found in the regressions. The reference category consists of home renters, who clearly had the highest child mortality. The lowest mortality group, with mortality levels 16 percent below those of home renters, comprises those owning farms free and clear. In general, the results for this variable show the advantages both of property ownership and of farm life. If one rented, it was better to rent a farm than a nonfarm home; if one owned, it was better to own a farm than a nonfarm home. So the advantages of farm life are not entirely captured by our rural/urban variable but show up as well in occupational results (children of farmers having below-average mortality) and in farm/nonfarm residence. The more bucolic was life on any of these dimensions, the lower was mortality.

Presence of Servants or Boarders

The presence of servants is expected to be negatively associated with child mortality because it is a sign of affluence. The presence of boarders is thought to be an indicator of economic hardship in the family (Modell and Hareven 1973). Further, attending to boarders might have taken the mother's time away from child care. Both servants and boarders furnished an additional vector for introducing infectious disease into the household and may have produced more crowded conditions. Tabulations presented in Table 3.1 were consistent with these general expectations, and the regression coefficients in Table 4.1 also support them. The signs are in the expected direction (negative for servants, positive for boarders), but only the coefficient for boarders is statistically significant. The presence of boarders raised child mortality by about 15 percent, roughly equivalent to the effect of having an unemployed husband, a far more common condition.

Migration Status

The 1900 census asked a question of migrants from abroad about the year of their immigration to the United States. From this datum and current age, two variables were computed, one giving the proportion of the first 14 years of a woman's life that were lived in the United States and the other giving the proportion of years lived after age 14 that were spent in the United States. For natives, the value of both variables is 1.00. Coefficients of these variables are very small and statistically insignificant, and their signs are opposite from one another in Table 4.1. They do not prove to be helpful in understanding child mortality variation, once the nativity of the mother is taken into account.

Multivariate Relations within the Native, Foreign-Born, Rural, and Urban Populations

The relations described above are based upon a model in which effects of different variables can be added together to produce a predicted level of child mortality. This model does not include the possibility of interactions among independent variables. For example, the effect of husband's literacy may be greater in urban than in rural areas, but the effect described in Table 4.1 is an average across both types of areas.

In order to examine certain of these interactions, we have reestimated equation (4) in Table 4.1 separately for four major groups: native-born women, foreign-born women, urban women, and rural women. It would also have been useful to examine relations separately for whites and blacks, but there are too few observations for, and too little internal differentiation within, the latter group to make such an exercise worthwhile.[3] The distinctions that we explore do not exhaust the possibilities of interactions, of course, but they do capture major divisions within the American population that are suspected of conditioning the relationship between child mortality and other variables.

Results for the four groups are presented in Table 4.4. Also shown for convenience is equation (4) for all groups combined, reproduced from Table 4.1.

Race

Table 3.2 demonstrated that the child mortality disadvantage of black women was much greater in urban than in rural areas. This result is repeated in Table 4.4 when other variables are controlled. The coefficient for black women is more than twice as large in urban as in rural areas, and both coefficients are highly significant. In urban areas, black mothers had 58 percent higher child mortality than white mothers. This is the largest coefficient appearing in Table 4.4 and demonstrates again why the predominantly urban black women in the Death Registration Area presented such a distorted picture of overall black mortality conditions.

TABLE 4.4
Equations Predicting the Ratio of Actual to Expected Child Deaths for Individual Women Distinguished by Nativity and Rural/Urban Residence: U.S., 1900

	Total		*Native-born*		*Foreign-born*		*Urban*		*Rural*	
Independent variables	*Coefficient*	*Significance*	*Coefficient*	*Significance*	*Coefficient*	*Significance*	*Coefficient*	*Significance*	*Coefficient*	*Significance*
Intercept	0.6280	***	0.6929	***	0.2131	***	0.3990	*	1.0262	***
Race of wife										
White	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI
Black	0.2968	***	0.2794	***	NI	NI	0.5796	***	0.2368	***
Nativity of wife										
Native with native mother	NI	NI	NI	NI	—	—	NI	NI	NI	NI
Native with British mother	−0.0345	—	−0.0314	—	—	—	−0.0108	—	−0.0609	—
Native with Irish mother	0.0562	—	0.0447	—	—	—	0.1743	**	−0.2687	**
Native with Scandinavian mother	−0.0664	—	−0.0737	—	—	—	−0.1106	—	−0.0732	—
Native with German mother	−0.0490	—	−0.0637	—	—	—	−0.0133	—	−0.0666	—
Native with other West European mother	−0.3055	*	−0.3041	*	—	—	−0.3715	—	−0.3001	—
Native with East European mother	−0.1479	—	−0.1909	—	—	—	−0.2657	—	−0.0391	—
Native with South European mother	−0.4353	—	−0.4305	—	—	—	−0.4017	—	−0.4312	—
British	0.0592	—	—	—	−0.0565	—	0.1486	—	−0.0239	—
Irish	0.1169	—	—	—	NI	NI	0.3023	*	−0.4069	—
Scandinavian	0.1290	—	—	—	0.0232	—	0.1053	—	0.0763	—
German	0.0923	—	—	—	−0.0186	—	0.2636	*	−0.1219	—
Other West European	−0.2089	—	—	—	−0.2899	*	−0.1472	—	−0.3296	—
East European	−0.0779	—	—	—	−0.1058	—	0.0488	—	−0.1985	—
South European	−0.0816	—	—	—	−0.1163	—	0.0275	—	−0.2501	—
Other-foreign born	0.0443	—	—	—	−0.1226	—	0.1991	—	−0.1204	—

Age of wife	0.0094	***	0.0056	**	0.0227	***	0.0123	***	0.0076	**
Literacy of wife										
Literate	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI
Illiterate	0.1004	**	0.1414	***	0.0036	—	0.0816	—	0.1166	**
Wife speaks English										
Yes	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI
No	0.0029	—	NI	NI	0.1123	—	0.0439	—	−0.0871	—
Literacy of husband										
Literate	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI
Illiterate	0.1234	***	0.1373	***	0.0977	—	0.2650	***	0.0766	—
Husband speaks English										
Yes	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI
No	0.1506	—	NI	NI	−0.0193	—	−0.0506	—	0.3637	**
Region										
New England	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI
Middle Atlantic	−0.0557	—	−0.0008	—	−0.1268	—	0.0441	—	−0.3618	**
East North Central	−0.0928	—	−0.0658	—	−0.1403	—	−0.0003	—	−0.3508	**
West North Central	−0.1302	**	−0.0822	—	−0.3010	**	−0.0828	—	−0.3792	**
South Atlantic	−0.2235	***	−0.2229	**	−0.1843	—	0.1436	—	−0.5777	***
East South Central	−0.1641	*	−0.1661	*	−0.2414	—	0.0738	—	−0.4812	***
West South Central	−0.0074	—	−0.0267	—	0.3875	*	0.2734	**	−0.3263	*
Mountain	0.4308	***	0.4094	***	0.4462	*	0.1718	—	0.2642	—
Pacific	−0.0615	—	−0.0090	—	−0.1758	—	−0.2565	*	−0.2225	—
Residence										
Top 10 cities	0.2954	***	0.3572	***	0.1951	*	0.2574	***	—	—
Other cities 25,000+	0.3160	***	0.3249	***	0.2594	**	0.2985	***	—	—
Cities 5,000–24,999	0.2040	***	0.1901	***	0.2181	*	0.1933	—	—	—
Cities 1,000–4,999	NI	NI	NI	NI	NI	NI	NI	NI	—	—
Rural	0.0682	—	0.0440	—	0.1614	—	—	—	—	—

TABLE 4.4 *(cont.)*

	Total		*Native-born*		*Foreign-born*		*Urban*		*Rural*	
Independent variables	*Coefficient*	*Significance*	*Coefficient*	*Significance*	*Coefficient*	*Significance*	*Coefficient*	*Significance*	*Coefficient*	*Significance*
Property										
Owns farm clear	−0.1577	**	−0.0872	—	−0.4119	**	−0.0757	—	−0.1568	**
Owns farm mortgaged	−0.1005	—	−0.0315	—	−0.3629	**	−0.3130	—	−0.0925	—
Rents farm	−0.0634	—	0.0002	—	−0.4240	**	−0.3071	—	−0.0293	—
Owns home clear	−0.0417	—	−0.0336	—	−0.0697	—	−0.0619	—	0.0119	—
Owns home mortgaged	−0.1064	**	−0.0828	—	−0.1728	**	−0.1005	*	−0.0947	—
Rents home	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI
Wife working										
Working—has child care	0.2139	—	0.2868	*	−0.5052	—	−0.1772	—	0.3436	*
Working—no child care	0.0357	—	−0.0132	—	0.2564	—	0.1506	—	−0.1039	—
Not working—has child care	0.0139	—	0.0117	—	0.0387	—	−0.0068	—	0.0335	—
Not working—no child care	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI
Migration status										
Proportion of life lived in the U.S. before age 14	−0.0266	—	NI	NI	0.0138	—	0.1294	—	−0.2304	—
Proportion of life lived in the U.S. after age 14	0.0186	—	NI	NI	0.1340	—	−0.0561	—	0.1054	—
Other										
Husband unemployed	0.1586	***	0.2188	***	0.0087	—	0.1227	**	0.1788	***
Servant(s) in household	−0.0211	—	0.0138	—	−0.2120	—	−0.1132	—	0.0662	—
Boarder(s) in household	0.1510	***	0.1245	**	0.2372	***	0.1384	**	0.1718	***
Income										
Husband's age	−0.0006	—	−0.0008	—	0.0027	—	−0.0026	—	0.0009	—
Husband's age squared	0.0000	—	0.0000	—	0.0000	—	0.0000	—	0.0000	—
Log of state income index	0.4851	***	0.5589	***	−0.2846	—	0.1289	—	−0.6646	***

Husband's occupation										
Professional, Technical	−0.0652	—	−0.0550	—	−0.1516	—	−0.0602	—	0.0353	—
Agricultural (excluding Laborers)	−0.1454	**	−0.1582	**	−0.1017	—	−0.0671	—	−0.0824	—
Agricultural Laborers	−0.0465	—	−0.0249	—	−0.1190	—	−0.1638	—	0.0624	—
Managers, Officials, Proprietors	−0.1098	*	−0.1258	*	−0.0808	—	−0.1315	*	−0.0150	—
Clerical	−0.1650	*	−0.1424	—	−0.2440	—	−0.2116	**	0.0875	—
Sales	−0.2093	**	−0.1898	*	−0.2335	—	−0.2330	**	−0.0539	—
Craftsmen, Foremen	−0.0217	—	−0.0099	—	−0.0726	—	−0.0355	—	0.0019	—
Operatives	−0.0904	*	−0.0324	—	−0.2442	**	−0.1370	**	0.0281	—
Service Workers	−0.1189	—	−0.0809	—	−0.2048	—	−0.2136	**	0.1521	—
Laborers	NI	NI	NI	NI	NI	NI	NI	NI	NI	NI
Miscellaneous	−0.1424	—	−0.1670	—	−0.0653	—	−0.1668	—	−0.0401	—
N	9,917		7,902		2,014		4,535		5,381	
Adjusted R-squared	0.0505		0.0515		0.0480		0.0405		0.0535	
F-ratio	8.875	***	8.661	***	2.752	***	3.902	***	5.905	***

Source: Sample of census enumerators' manuscripts, U.S., 1900.

Note: Sample consists of currently married women with husband present, married less than 25 years. The dependent variable is the ratio of actual to expected child deaths for each woman. For an explanation of its construction, and of the other variables, see text. Regressions are weighted least squares with the weights being the number of children ever born (normalized back to the original number of observations).

Significance levels are: *** = significant at least at a 1 percent level; ** = significant at least at a 5 percent level; * = significant at least at a 10 percent level; — = not significant at least at a 10 percent level. NI = not included.

Nativity and Parentage

Once other factors are controlled, second-generation mothers continue to have unusually low mortality, whether the sample is limited to all native-born women, to rural women, or to urban women. Nineteen of the 21 coefficients for this group in Table 4.4 are negative, signifying child mortality levels below those of native white women of native parentage. Only the urban and rural coefficients for second-generation Irish women are significant, and these have opposite signs: women of Irish heritage did poorly in cities and quite well in rural areas. Note that this result also pertains to first-generation immigrants from Ireland, although only the coefficient for urban Irish women attained statistical significance. It is likely that migrant selectivity factors are responsible for this unusual pattern of results for Irish women. It has been suggested that rural Irish in the United States were more likely to have been Protestant and to have had some wealth when they left Ireland than the urban Irish.[4] In addition, migrants from overwhelmingly rural Ireland who settled in urban areas undoubtedly faced larger problems of adaptation than did those settling in rural areas.

Foreign-born women in general did well in rural areas and poorly in urban areas, although none of the coefficients is significant at a 5 percent level. Nevertheless, the two largest immigrant groups, German- and Irish-born women, had child mortality levels in urban areas that were 26–30 percent above those of native women with native parents, coefficients that are significant at a 10 percent level.

The literature of the era is replete with references to the crowded, dilapidated, substandard conditions in immigrant ghettos. For example, the noted social reformer Jane Addams wrote of Chicago:

> The most obvious faults were those connected with the congested housing of the immigrant population, nine-tenths of them from the country, who carried on all sorts of traditional activities in the crowded tenements. That a group of Greeks should be allowed to slaughter sheep in a basement, that Italian women should be allowed to sort over rags collected from the city dumps, not only within the city limits but in a court swarming with little children, that immigrant bakers should continue unmolested to bake bread for their neighbors in unspeakably filthy spaces under the pavement, appeared incredible to visitors accustomed to careful city regulation. . . .
>
> Many evils constantly arise in Chicago from congested housing. . . . The inevitable boarders crowded into a dark tenement already too small for the use of the immigrant family occupying it . . . The school children

> who cannot find a quiet spot in which to read or study and who perforce go into the streets each evening; the tuberculosis superinduced and fostered by the inadequate rooms and breathing spaces. (Addams 1910:294, 296)

Addams reported that, during the typhoid epidemic in Chicago in the summer of 1902, approximately one-sixth of the deaths arose in one ward, which contained only one-thirty-sixth of the population. Much of the blame was attributed to poor sanitary facilities (Addams 1910:296–99).

Another noted publicist and reformer, Jacob Riis, wrote in his famous books *How the Other Half Lives* (1890) and *The Children of the Poor* (1892) of such notorious New York City neighborhoods as Hell's Kitchen, the Five Points, the Bend, and Battle Row. His photographs were equally famous. One of his descriptions concerns the infamous Gotham Court, largely populated by Irish and Italians, which had a death rate of 195 per 1000 during the last large cholera epidemic in New York City (1866). An infant mortality rate of 442 per 1000 live births had been reported in the 1860s (Riis 1890:35–36). The descriptions included common elements: small rooms, crowded conditions, poor heat in winter, few windows, deficient ventilation, and inadequate water and sewerage disposal. Oscar Handlin, in his work on immigrants to the United States, *The Uprooted* (1973), uses the technique of the "ideal type" to describe the immigrant ghetto. Some of his most lurid passages deal with the congestion of residences and the filth to be found both within the tenements and in the streets below (e.g., Handlin 1973:135–36).

Such accounts tended to focus attention on the most extreme circumstances rather than on the most typical. Immigrants had elevated mortality rates in cities, but so did natives. According to Table 4.4, native-born women living in the 10 largest cities suffered 36 percent higher child mortality than those living in cities with populations of 1,000–4,999. Apart from German and Irish immigrants, coefficients for foreign-born urban mothers in Table 4.4, though generally positive, are insignificant. Likewise, Higgs and Booth (1979) also found slight effects of nativity on mortality in the 1890s once other factors were controlled. Whatever extra hazards were present in the first generation had disappeared for all but the Irish in the second generation. Their risks pale in comparison to the 47 percent excess for the black population (from Table 3.1), very few of whom faced the stresses of tenement living. By 1900, the melting pot was already percolating, though blacks had yet to gain entrance.

7. The congested conditions of urban tenements, such as that pictured here in New York (ca. 1910), were undoubtedly related to the excessive child mortality of both natives and immigrants in cities.

Age of Mother

Consistent with earlier results, mortality was declining slightly faster for urban than for rural areas (i.e., the coefficient of mother's age was greater in the former), although the controls for other characteristics that are instituted in Table 4.4 seem to reduce the rural/urban disparity. More impressively, the foreign born had a rate-of-mortality decline that was more than three times greater than the rate among natives. This disparity may reflect in part the difficulties that some of the foreign-born women faced in their country of origin or en route to the United States, although we have controlled the length of time that they had spent in the U.S.

Literacy and English-Speaking Ability

The urban and rural coefficients for mother's literacy in Table 4.4 are similar in magnitude, though only the rural is statistically significant. Literacy did have a higher payoff among native than among immigrant women, perhaps because the census question referred to literacy in any language. Among immigrant women, ability to speak English was more important than literacy, with a coefficient similar to that of literacy among native women.

Husband's literacy and English-speaking ability are far more differentiated in their effects. Literacy was highly significant in urban areas and had more than three times the influence that it had in rural areas. It is likely that husband's literacy had a higher economic value in urban areas. It was also three times as consequential there as the wife's literacy, perhaps because the wife was rarely in the labor force. A United Nations (1985) study of 15 developing countries also found that husband's literacy makes a bigger difference for child mortality in urban than in rural areas. The husband's literacy in 1900 had smaller effects for child mortality among immigrants than among natives-born wives.

The surprising result for the husband's English-speaking ability is that it was far more consequential in rural areas than in urban. Perhaps this difference reflected a shortage of other people who could speak a foreign language and act as intermediaries with supervisors and foremen in rural areas. It may also reflect the fact that spoken discourse was more important in rural areas and written discourse in urban. These explanations may be too facile, but there is no denying the very different pattern of results in urban and rural areas: in urban areas, husband's literacy mattered a great deal for child survival and his English-speaking ability was essentially irrelevant, whereas just the opposite relations prevailed in rural areas.

Region and Size of Place

The relatively high mortality of New England, the reference category for regions of the U.S., is largely traceable to rural areas. Only rural residents of the Mountain states had worse mortality. Soil quality is relatively poor in much of New England, and since it was the earliest settled region, its rural regions were unusually densely populated. The ratio of *rural* population to land area was at least three times the national average in Connecticut, Massachusetts, and Rhode Island (U.S. Bureau of the Census 1975: Series A 195–209 and A 210–263). New England was also the most urbanized region in 1900, however, so that our sample contains only 81 women and 214 children ever born in rural New England (8.3 percent and 8.5 percent of New England observations, respectively). It is possible that this small sample has given a misleading impression of rural conditions there.

Rural South Atlantic areas had 58 percent lower mortality than rural New England once other factors were controlled, and rural areas of East South Central states were also exceptionally well off. The Mountain region is poorly situated on all regressions, while the advantages of the Midwest (East North Central and West North Central) are seen to be concentrated in rural areas and, to a lesser extent, among the foreign born. Adding to the advantages of the Midwest and South in overall rankings was their predominantly rural character.

Large cities were clearly harmful for both immigrants and natives. For each group distinguished, mortality was 20–36 percent greater in the two largest city categories than in cities of 1,000–4,999 inhabitants. The curvilinear pattern that was evident in the overall result, with the top 10 cities doing better than the next largest group, was confined to the foreign born, as suggested earlier in Table 3.2.

Economic Circumstances

We expected to find that child mortality would be highest for working mothers without people in the household who might be available for child care. That result was not found in the total sample. This result does emerge in urban areas, however, although it is not statistically significant. Women's work, and attendant child-care concerns, appear more consequential in urban areas, possibly because more of the work occurred outside of the home.

Having an unemployed husband induced positive, relatively large, and statistically significant effects in all the equations in Table 4.4 except among the foreign born. For the native population, and in both rural and urban areas, a period of unemployment by the principal income earner in the family during the census year raised the child mortality index by 12–22 percent. The coefficient is largest for the native population.

For the foreign born and in urban areas, the presence of one or more live-in servants was associated (though not significantly) with lower child mortality, while among the native born and in rural areas the effect was smaller and opposite in direction. In contrast, the presence of boarders in the household had a uniformly unfavorable and significant impact on child survival in all of the equations of Table 4.4. As mentioned earlier, a number of factors may be involved in this outcome: boarders as an indicator of economic distress in the household, as a potential source of infection, and as an additional burden on a mother's time. The effect of boarders on child mortality was about equally strong in urban and rural areas, but it was twice as strong among the foreign born as among native-born women, perhaps because the foreign born were more disadvantaged to begin with.

The final set of variables relates to income and occupation. A curvilinear relation between child mortality and husband's age was anticipated because of the curvilinear pattern of male income as a function of age, but such a pattern failed to appear in the full sample or among any of the subgroups.

State earnings levels were highly significant in the total sample. They also proved significant in rural areas and among the native population, where coefficients are even larger than in the total sample. Since both groups contained a majority of the population, they contributed the most observations to the construction of the earnings index, and it is perhaps not surprising that its variation is most clearly reflected therein. But we also expected significant and sizable coefficients in urban areas and among the foreign born, where they fail to appear. In any event, state earnings levels appear to be a good reflection of average conditions in rural areas, and to be closely associated with child mortality levels therein.

Breaking the sample down into urban and rural components sheds considerable light on the role of husband's occupation. The effects of occupation on child mortality appear greater in urban areas than in rural areas or in the country as a whole. Interoccupational mortality variation was greater, and significant coefficients more common. Relative to the low-status reference category (urban laborers), urban clerical workers, sales personnel, operatives, and service workers had significantly lower mortality. Professionals continued to do poorly in urban areas, however. Occupational mortality differentials were smaller in rural areas, and none of the rural groups had mortality that differed significantly from that of rural laborers. Occupational differentials among the foreign born, who were predominantly urban, were in general quite similar to those in urban areas as a whole.

It is possible that occupation was more closely related to income in urban areas than it was in rural areas. This situation might have arisen simply because occupational labelling is a more precise process in the more highly differentiated urban industrial establishments. Furthermore, rural workers may have changed jobs more frequently and casually, partly in response to the rhythms of the crop cycle. A less differentiated character of occupational pursuits in rural areas may partly explain why the state income level works so much more powerfully there. In any event, the relatively unstructured relations between occupation and child mortality in rural areas are reflected in the total sample, within which rural residents formed a majority. But it is reassuring to find that our expectations about occupational mortality differences—based as they inevitably are on experience in modern urbanized populations—are somewhat more clearly realized in urban areas.

The Relative Importance of Variables

The previous sections have described the degree to which child mortality was differentiated according to various characteristics, once other variables were controlled. To some extent, we were able to elucidate the paths through which characteristics that are established early affected mortality, and we tested the significance of differences between mortality levels within various categories of a variable relative to the reference category.

While these tests are illuminating, they do not tell us everything we want to know. They indicate how child mortality varied from one category of a variable to another, but not how important the *set* of categories (i.e., the variable) was relative to other sets. Of course, if there was little or no variation in mortality levels within the set, then the variable itself cannot be very helpful in explaining variation in child mortality.

To shed additional light on the issue of how important one variable was relative to others, we ask two specific questions:

1. If one knew only one piece of information about a family (e.g., husband's occupation, wife's literacy), which piece would be most helpful in predicting the family's level of child mortality?
2. If one piece of information about a family were missing, which omission would be most costly in terms of predicting the family's level of child mortality?

These questions can be translated into relatively simple statistical terms. The first question amounts to asking how much variance in child mortality can be explained by one variable acting alone, whereas the second question amounts to asking how much of the variation explained is lost when one variable is deleted from a model with all other variables included.

We would expect that variables with more categories would explain more variance even if child mortality were a purely random process, simply because more categories provide more opportunities for a spurious association to emerge. So it is useful to introduce tests of statistical significance that account for the greater degrees of freedom used up by variables with more categories.

Table 4.5 presents the appropriate statistics. The answer to the first question is very surprising. The single most valuable piece of information for predicting child mortality is the mother's race, despite the fact that the variable is a dichotomy.[5] That is to say, the value of R^2 for this variable, shown in the first column of Table 4.5, exceeds that of any other variable. The next most valuable pieces of information are size of place of residence, husband's occupation, and farm/home property ownership. But these variables have many categories, especially the first two, and when an adjustment is made for degrees of freedom, as for the F-ratio, they drop down in the hierarchy of variables. After race, the three variables with the highest F-ratios are all dichotomies: whether the husband was unemployed and whether the husband or wife was illiterate. All of the variables, however, when considered individually, are significantly related to mortality at a 1 percent level. Were we forced to consider variables one at a time, it would appear that child mortality was powerfully influenced by every factor that we can measure.

Multivariate analysis proves more discriminating. Fewer than half of the variables contribute significantly at a 1 percent level to explaining variation in child mortality when all other variables are present in the model. Once again, however, the most important variable turns out to be race. Adding race to the model increases the variance explained by .0052, or by 10 percent. Being black in 1900 denoted a set of economic and social conditions that powerfully affected child mortality and that was not adequately captured by other variables that may have been associated with race. The second most important variable is size of place, adding .0047 to variance explained. Larger cities were unhealthier, and the population was sufficiently dispersed by size of place that the variable explains a substantial amount of variance. The third most important variable was region of residence (controlling state income, among other factors), pointing again to the potential importance of environmental conditions as a factor in mortality variation. These three variables add at least twice as much explanatory power as any other variables in the model.

All three remain highly significant after accounting for degrees of freedom, but the multicategory size-of-place and region-of-residence variables are surpassed in F-ratios by more parsimonious variables: husband's unemployment, state income levels, presence of boarders in the household, and wife's age.[6]

Despite having many categories, wife's nativity and husband's occupation add only .0017 and .0014 to variance explained, amounts that are statistically insignificant. The literacy status of the wife is also insignificant at a 5 percent level. If we knew everything else about a family, we would gain no significant additional information about its

TABLE 4.5

Statistical Significance of Variables in the Multivariate Analysis of Child Mortality: U.S., 1900

	Variables Alone				Full Model				
Variable	*R-square*	*Adjusted R-square*	*F-ratio*	*Prob>F*	*R-square without predictor*	*Adjusted R-square without predictor*	*Marginal R-square*	*F-ratio*	*Significance*[a]
Race of wife	0.0150	0.0148	75.392	0.0001	0.0517	0.0454	0.0052	27.152	***
Nativity of wife	0.0064	0.0048	3.992	0.0001	0.0552	0.0503	0.0017	1.110	—
Age of wife	0.0015	0.0014	14.924	0.0001	0.0556	0.0493	0.0013	13.576	***
Literacy of wife	0.0096	0.0094	48.158	0.0001	0.0564	0.0501	0.0005	2.611	*
Wife speaks English	0.0024	0.0022	11.820	0.0001	0.0569	0.0507	0.0000	0.000	—
Literacy of husband	0.0086	0.0084	42.799	0.0001	0.0562	0.0499	0.0007	3.655	**
Husband speaks English	0.0027	0.0026	13.615	0.0001	0.0567	0.0505	0.0002	1.044	—
Region	0.0082	0.0074	10.256	0.0001	0.0529	0.0472	0.0040	5.222	***
Residence	0.0116	0.0111	23.239	0.0001	0.0522	0.0463	0.0047	9.817	***
Property ownership	0.0134	0.0128	22.404	0.0001	0.0557	0.0498	0.0012	2.089	*
Wife working	0.0029	0.0026	9.683	0.0001	0.0567	0.0506	0.0002	0.696	—
Husband unemployed	0.0064	0.0063	63.618	0.0001	0.0549	0.0486	0.0020	20.886	***

Servants in household	0.0010	0.0009	9.663	0.0019	0.0569	0.0506	0.0000	0.000	—
Boarders in household	0.0018	0.0017	18.079	0.0001	0.0557	0.0494	0.0012	12.532	***
Log state income index	0.0022	0.0021	22.272	0.0001	0.0553	0.0489	0.0016	16.709	***
Husband's occupation	0.0129	0.0119	12.955	0.0001	0.0555	0.0500	0.0014	1.462	—
All variables plus covariates[b]	0.0569	0.0505	8.875	0.0001					

Source: For sources and methods, see text.

Note: N = 9,917. Prob>F is the probability of observing the value of the F-ratio if there were no relationship between the variable and mortality.

[a] Significance levels are: *** = significant at least at a 1 percent level; ** = significant at least at a 5 percent level; * = significant at least at a 10 percent level; — = not significant at least at a 10 percent level.

[b] The covariates are age and age squared of husband and two variables on the proportion of life spent in the United States.

child mortality levels by knowing the husband's job or whether the wife was literate or what her ethnic background was. Since these variables are prominently featured in current discussions of child mortality (see Chapter 5 and United Nations 1985), these results are quite surprising.

A word of caution is necessary. As we argued earlier, both literacy and ethnic background to some extent work through other variables whose values are established later in life and which are controlled in the model. This layering of influence may also be operative to some extent for husband's occupation. All that these results convey is that little *additional* variance is explained when these variables are introduced; in the parlance of path analysis, no significant *direct* effect of these variables is evident. Furthermore, some of the explained variance is shared among variables, and the test reported in Table 4.5 allows all of that shared variance to be captured by the other variables that remain in the equation when individual variables are excluded. For example, in Table 4.1, the literacy of the wife remains a significant variable when all other variables are present because it appropriates some fraction of the explained variance that it shares with other variables. Nevertheless, the test in Table 4.5 is an appropriate way of answering the second question posed: we would lose nothing of significance in explaining child mortality variation in 1900 if we had no information on the wife's literacy or ethnicity or on her husband's occupation.

Taking variables one at a time as we have just done does not allow for the fact that several variables may be tapping into a common root. A final way to characterize variables, one that permits some cumulation of influences, is to list them according to the largest contrasts that occur between two categories within the variable, once other variables are controlled. In order of the size of mortality differences between two categories, and limiting ourselves to categories that accounted for at least 5 percent of births in the sample, the list is shown at the top of the next page (from equation [4] of Table 4.1).

Three of these variables (occupation, home/farm ownership, and size of place) are related to rurality, a pervasive influence on mortality at the time. Assuming that the influence of variables is additive, a laborer living in a city larger than 25,000 inhabitants who rented his home had mortality that was 61 percent higher than that of a farmer in a small town who owned his farm (.15 + .30 + .16). Clearly, farm living, farming occupations, and rural residence all contributed independently to the health advantages of a more pastoral life.

What about economic variables as a class? The surprising result is that little occupational variation existed in child mortality, whether

High-mortality category	*Low-mortality category*	*Difference in mortality*
Black	White	.30
City > 25,000	Town < 5,000	.30
New England	South Atlantic	.22
Husband unemployed	Husband not unemployed	.16
Rents home	Owns farm clear	.16
Boarders in household	No boarders	.15
Laborer	Farmer	.15
Husband illiterate	Husband literate	.12
Wife illiterate	Wife literate	.10
Wife speaks no English	Wife speaks English	.00

we use a set of occupational categories or indexes of occupational income and unemployment. Still, laborers had 19 percent higher mortality than clerical and sales workers, the low-mortality white-collar groups. And if we add the unemployment penalty of 16 percent and place the clerical/sales worker in a state with double the per capita income of the laborer's state, adding a mortality differential of (*ln* 2 x .485 = .34), we generate a contrast between rich and poor of 69 percent. Economic discrepancies are clearly associated with large variation in child mortality, although no single economic variable shows striking effects by itself. In addition, the large mortality variation by race that we have demonstrated is most plausibly ascribed to the enormous economic disparities that existed between the races at the time.

The variables that do not appear to be very important in child mortality, individually or as a group, are those which we expect to be most closely associated with child-care practices: mother's literacy, her ethnicity, her English-speaking ability, and her husband's occupation. None of these variables significantly explains variation at a 5 percent level in Table 4.5, once other variables are controlled. Whatever behavioral variation was associated with these variables seems to be swamped in its effects by broad geographic and economic factors. The seven variables that are significant at a 1 percent level in Table 4.5 include three geographic variables—size of place, region of residence, and state income levels—and three that denote individual economic circumstances—race, husband's unemployment, and presence of boarders. The seventh is the woman's age, which represents the health environment at the time when her children were young.

This list paints individuals as relatively passive victims of time, place, and labor markets. Behavioral factors were probably a key to child mortality declines during the twentieth century (Ewbank and

Preston 1989), but their mark was not highly visible at the turn of the century in variables such as literacy and occupation where we might expect their effects to be most obvious. One reason may be that too little knowledge of specific ways to enhance child survival had developed to allow individuals to escape from the circumstances imposed by broad geographic and economic forces. We develop this theme at greater length in the next chapter.