This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

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: Notes

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

Chapter URL: http://www.nber.org/chapters/c11550

Chapter pages in book: (p. 229 - 236)

NOTES

CHAPTER ONE

1. This result refers to the surviving-children method using the West life table. (See Chapter 2.)

2. The states were Connecticut, Massachusetts, New Hampshire, New Jersey, New York, Rhode Island, and Vermont.

3. Industrial structure was defined as the proportion of males aged 10 and over employed in mines and quarries; in metals, etc.; in woodworking, etc.; in brickmaking, etc.; in skins and hides, etc.; in textiles; and as mechanics and laborers.

4. Evidence from England and Wales in 1911 also reveals a monotonic inverse relationship between child mortality and number of rooms per family (see Table 5.1).

CHAPTER TWO

1. The Death Registration Area was formed in 1900 from states and cities where tests revealed that death registration was at least 90 percent complete and where a uniform death certificate had been adopted (Cassedy 1965). It consisted originally of Maine, Vermont, New Hampshire, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, Michigan, Indiana, the District of Columbia, and selected cities not in these jurisdictions. In this chapter, numerical results for the Death Registration Area refer to the 10 states just enumerated and the District of Columbia, collectively referred to by Glover (1921) as the "Original Registration States."

2. The mortality statistics for the Death Registration Area itself reveal considerably higher mortality in urban areas and among blacks (Glover 1921).

3. Comparable questions on children ever born and children surviving were asked in the 1890 census, but unfortunately most of those original schedules were lost in a fire in 1921. The questions were asked again in the 1910 census, but the results were not then tabulated or used. A project was finally undertaken in conjunction with the 1940 census that retabulated and analyzed the 1910 data for children ever born (U.S. Bureau of the Census 1943, 1944, 1945). But the information on children surviving was never tabulated or used.

4. Comparisons of various age, sex, race, and nativity distributions with published data using a series of chi-square tests revealed generally statistically insignificant differences.

5. This "woman" file contained the census information on the woman herself, that of her husband and matched own-children when present, and selected characteristics of the area of residence. The actual matching procedure was carried out at the University of Washington, under the supervision

of Avery Guest and Stephen Graham. We are grateful to them for making this file available.

6. For additional details on these procedures and on this methodology, see Preston and Palloni 1978. An application to 1900 census data for the anthracite mining region of Pennsylvania is made in Haines 1977.

7. This is based on the formula that the variance of a binomial distribution is $p \times q \times n$, where p = the probability of the outcome occurring (i.e., in this case, a child death); $q = (1 - p)$; and n = the number of cases (or trials). The standard deviation is the square root of the variance.

8. Because of the small number of children ever born in the census sample for several states and territories, some had to be combined. Thus Arizona, Nevada, and New Mexico were combined into one unit, as were Idaho, Montana, and Wyoming. Oklahoma and Indian Territory were taken as one unit. The District of Columbia was used as a unit of observation. Alaska and Hawaii were excluded. There are thus 45 observations in total.

9. Among the states with positive residuals:

State	*Residual*	*Registration coverage in the 1900 census (percent)*
Pennsylvania	4.43	41.1
Ohio	2.48	31.9
Maryland	3.85	44.3
Delaware	11.99	41.4
California	0.16	41.1
Wisconsin	0.70	20.9

Registration coverage is from Condran and Crimmins 1979: Table 2.

CHAPTER THREE

1. The male and female life tables for this level of mortality were combined by weighting the $1(x)$ column of the life tables by 0.5122 for males and 0.4878 for females on the assumption of a sex ratio at birth of 105 males per 100 females.

2. An example would be a case in which mortality has been declining in the past and in which the intention is to compare the wives of farmers to those of farm laborers. If farmers and their wives were, on average, older and of higher marital duration than farm laborers and their wives, then the index of child mortalty for farmers will tend to be somewhat overstated relative to that for farm laborers. Since farmers tended to have lower child mortality, the effect in this instance would be to decrease the differential.

3. Note that, to estimate the significance of a difference in child mortality between two groups, the t-test is

$$t = \frac{q_A(5) - q_B(5)}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}}$$

where s_i^2 is the sample variance of $q(5)$ in group i and n_i is the number of births in group i.

4. Note that in this chapter we use the 1900 census definition of an urban place, which has a minimum size of 1,000 inhabitants. Several tabulations in earlier chapters used a size criterion of 2,500 for the purpose of establishing comparability over time.

5. The enumerators' instructions were to answer "Yes" in the "Can Read" column for any person aged 10 and over who could read in any language. A "No" was entered otherwise. All children below age 10 were to be given a blank. A similar instruction was made for the column "Can Write." "The inquiries . . . are intended to show the literacy of all persons 10 years of age and over, and should be answered according as they are able to read or write the language ordinarily spoken by them" (U.S. Bureau of the Census 1979:37). We adopt a definition of literacy requiring the ability to both read and write.

6. The proportions of illiterate women by race and ethnicity in the total sample population were: total, 11.9 percent; white, 7.1 percent; native white, 5.0 percent; native white with native mother, 5.7 percent; native white with foreign-born mother, 2.3 percent; foreign-born white, 14.8 percent; black, 51.4 percent.

7. Of the total households in the 1900 census sample (27,069 in number), 84.4 percent contained no extended family and 62.2 percent were husband/wife units (with or without children).

CHAPTER FOUR

1. Race is not included for the husband since, in the sample, there were almost no cases of racial intermarriage and hence race of husband and wife are nearly perfectly collinear.

2. Although state real income may be conceptually superior to state nominal income, the adjustment for price levels introduces additional error into the variables. Perhaps for this reason, the F-ratio for the equation using nominal income (8.875) exceeds that for the equation using real income (8.813). Consequently, we use nominal income for the basic presentations in this chapter. Some of the regional mortality differences that remain when nominal income is controlled will, therefore, reflect differences in price levels among the regions.

3. George Alter raised with us the possibility that results we presented regarding region (especially the low mortality levels in the South Atlantic) and working wives may be confined to the black population. To test this possibility, we reran equation (4) of Table 4.1 for whites alone. Results are not appreciably altered on these or other variables, although the mortality

advantage of the *rural* South is smaller when results are confined to the white population.

4. We wish to thank Timothy Guinnane for this suggestion.

5. Actually, a very small fraction of the population was Asian or had an unknown race, so three categories of this variable are recognized in the statistical analysis.

6. The F-ratio in this case is a measure of the statistical significance of the change in R^2 caused by removing a variable from the full model.

CHAPTER FIVE

1. The singulate mean age at first marriage was calculated from data on age, sex, and marital status from the 1911 Census of England and Wales, according to Hajnal's method (Hajnal 1953).

2. These categories are based on relatively "rough and ready" occupational groupings of the Registrar General, T.H.C. Stevenson. For a critique, see Armstrong 1972:203–6 and Szreter 1984. An extended treatment of childhood mortality by occupation of parents is given in Haines 1985.

3. The six county boroughs for Ireland were Dublin, Belfast, Cork, Londonderry, Limerick, and Waterford.

4. This result uses the mortality of the six county boroughs as the urban child mortality level and of the rest of Ireland as the rural figure. The standardization to England's urban percentage is thus .78 (.2058) + .22 (.1226) = .187.

5. As a check on the data for England and Wales, the differentials from the census estimates were compared to those from the first vital statistics tabulation (1911) of the infant mortality rate by the eight social-class categories used by the Registrar General at that time (Great Britain 1913b:73, 88). The results are as follows (indexed to a total of 100):

Social class	*Census*	*Vital statistics*
Professional, Higher White-Collar	65	61
Farmers & Lower White-Collar	83	85
Skilled Manual Workers	95	90
Semi-skilled Manual Workers	100	97
Unskilled Manual Workers	122	122
Textile Workers	120	119
Miners	128	128
Agricultural Workers & Related	79	78
Total	100	100

The two sets of results are highly correlated, despite the fact that the census results apply to children as well as infants, while the vital statistics apply only to infants. Further, the vital statistics data come from only one year of experience (1911) and are thus subject to more variability. The results apply to slightly different periods as well: the vital statistics to 1911 and the census index to about 1907–8. In addition, the 1911 census results apply to current

occupation, which may have differed from occupation in the past. The zero-order correlation between the mortality index for 116 detailed occupational categories in the census and the infant mortality rate for the same categories from the 1911 vital statistics indicates a quite close relationship ($r = 0.939$). Overall, the census mortality index seems to give an excellent picture of relative mortality by occupation of father for England and Wales just after the turn of the century.

6. Income measures are not strictly comparable between the United States and England. American incomes were largely yearly incomes from the 1901 Commissioner of Labor Survey, while English incomes were full-time equivalent annual earnings. Nonetheless, the relative incomes, as given by the index, should be independent of level. The index reflects higher relative incomes among most English professionals. A conversion rate of $4.85 per pound sterling was used, the prevailing rate under the gold standard. The difference in child mortality levels between England and the United States for specific occupational groups was tested for statistical significance. To do this, several steps were necessary. First, the index values were normalized so that the national average values were 1.0000. Second, these normalized values were converted to $q(5)$ values by multiplying them by the $q(5)$ value in the standard American model life table (West model, level 13, both sexes combined assuming a sex ratio at birth of 1.05; $q(5) = .19119$). This basically reduced both sets of differentials to a common mortality *level*. Third, it was assumed that these $q(5)$ values approximately followed a binomial process where $q(5) = p$ and $\delta = [p(1 - p)/n]^{1/2}$, where n was the number of children ever born to that occupational group. Finally, the statistical difference of these values was tested using the formula

$$t = (p_{us} - p_{ew})/(\delta^2_{us} + \delta^2_{ew})^{1/2}.$$

This yielded the following t-statistics:

Textile workers	.177
Miners	4.559
Agricultural laborers	
(a) Total	6.629
(b) U.S. native white	1.569
Teachers	2.582
Clergy	2.236
Clerks (commercial and business)	2.702
Physicians and surgeons	3.468

All but textile workers and native white agricultural laborers showed differences that were significant at least at a five percent level (two-tailed test).

7. For the United States, each of the occupational codes, based on the 1900 census coding system, was assigned an average annual income. Again, farmers are excluded because it was felt that the earnings of agricultural laborers (for whom an estimate was made) were not representative of farmers, and no good way could be found to capture the effect of differences in returns

from capital and land, a major component of farm incomes. The dates covered were in the range 1899–1901. (See Appendix A.) All estimates for grouped data were aggregated up from the sample.

For England and Wales, the period covered was 1901–14, but most of the estimates applied to 1906. The British earnings data come mostly as weekly, not average annual, earnings. In view of the difficulty of determining differences in weeks worked, it was decided to calculate full-time equivalent earnings by multiplying the average weekly earnings by 52, the practice of two of the sources (Williamson 1982b:36–37; Routh 1965:60–61). Consequently, the results for Britain are not comparable in level to those for the United States, although the years 1899–1901 were years of fairly low unemployment in the United States (U.S. Bureau of the Census 1975: Series D 85–86). Data were taken for England, or England and Wales, rather than for Great Britain whenever possible. (A number of the estimates included data for Scotland, but the results are dominated by the English data.) The estimates were made for 116 occupations or occupational groups. Any reaggregations—of social classes, for instance—were from these 116 groups. Groups for whom estimates were not available were farmers and a number of proprietors, such as retail merchants. Most professional and manual workers were covered, constituting about 75 percent of the male labor force in England and Wales.

8. There are small discrepancies in the occupations that could be used to estimate earnings inequalities, so that the result is not definitive. In particular, the absence of data for "proprietors" in England and Wales might have biased downwards the income level of the lower white-collar group, although the effect on measures of overall inequality is unclear. The groups actually included in the category for England and Wales are: Farmer's Sons; Officers of Local Authorities; Goldsmiths and Silversmiths; Watchmakers and Clockmakers; Bakers and Confectioners. Missing were such groups as: Coffee and Eating House Keepers; Inn and Hotel Keepers; Publicans; Boarding and Lodging House Keepers; Dealers and Merchants in various products (for example, coal, timber, wood, cork, bark, boots and shoes, corn, flour, seed); Drapers; General Shopkeepers; Grocers; Greengrocers; Tobacconists; Milksellers and Dairymen; Cheesemongers and Buttermen; Fishmongers, Poulterers and Game Dealers; Clothiers and Outfitters.

9. Using a 5000-inhabitant cut-off to define an urban place in the U.S. makes the U.S. definition as close as possible to that in England. The urban proportion in our sample for the U.S. is .37, and in England it was .73. The variance in proportion urban is $U(1 - U)$, where U is the percentage living in urban places. The variance is thus .172 in England and Wales and .233 in the U.S. The differences are not large, and the fact that the variance is greater in the U.S. offsets the fact that the sensitivity to urban residence was greater in England.

10. To clarify a bit more formally, suppose that mortality is a simple linear function of two variables:

$$M_i = a + bY_i + cU_i.$$

Then it can be shown that the variance in mortality is equal to

$$\sigma_M^2 = b^2\sigma_Y^2 + c^2\sigma_U^2 + 2bc\mathrm{COV}_{Y,U}.$$

The difference between values of σ_M^2 for two countries can be decomposed into a series of terms involving coefficients and variances of independent variables. The basic terms involving the variance in urban proportions (U) will disappear if δ_U^2 is the same in both countries. Also, since b is nearly the same in the two countries, the contribution of differences in δ_Y^2 will simply be weighted by b^2. We don't implement this scheme formally because of nonlinearities and differences in measurement in the two countries, but the logic underlies our analysis.

11. For an additional example, when variables representing the percentages of various occupations that are black, have literate wives, and have foreign-born wives are added to Model 2 for the United States, only the percentage of blacks is significant, with a coefficient of 10.1.

12. The total number of "expected deaths" among offspring of the intelligentsia is 130.2, apportioned in the following way: physicians, surgeons, and dentists, 35.6; clergymen, 32.3; teachers, 26.0; lawyers, 22.4; and pharmacists, 13.9.