Effect of early-life exposure to water-borne diseases on old-age survival in the United States

Magdalena Muszyńska* Seth Sanders †

PRELIMINARY

April 17, 2014

Abstract

We investigate whether exposure in early life to water-borne diseases leads to permanent scarring and decreases in old-age survival. Given the research hypothesis holds true, cohorts born in large American cities immediately after water treatment was widely implemented would have lower mortality at old age than cohorts born before. We estimate a set linear probability models using data for Americans 75 years and older in 1975-2001 from the Social Security Medicare Part B merged to the NUMIDENT files. We demonstrate that for those born at the beginning of the 20th century, cohorts born after a water treatment was introduced in a city of birth had higher rates of survival at old age than those born prior to this health intervention.

*Warsaw School of Economics, Poland
†Duke University, United States

Introduction

A large part of the urban penalty in mortality at the beginning of the 20th century was constituted by the prevalence of water-borne diseases. Access to clean drinking water in the American cities was in large part responsible for reductions in waterborne diseases.[1] Altogether, introduction of water treatment systems is estimated to be responsible for about half of the mortality reduction in the United States between 1900 and 1930 (Cutler et al., 2006; Cutler and Grant, 2005). These estimates refer to a short-term and mid-term effects of improved access to clean water. In the short-term, access to clean drinking water reduced the risk of waterborne diseases (i.e. typhoid fever) and related infectious diseases (i.e. diarrhea deaths among infants and very young children). In the mid-term, exposure to waterborne diseases can weaken the immune system and makes death from infectious diseases more lkely (i.e. pneumonia, tuberculosis) (Clay and Troesken, 2006; Cutler and Grant, 2005; Ferrie and Troesken, 2008; Sedgwick and MacNutt, 1910; Troesken, 2004).[2] In the long-term, for exam- ple, it has been demonstrated that typhoid carriers are more likely to develop cancer of the gallbladder, pancreas, colorectum, lung cancers as well as other malignant neoplasms (Caygill et al., 1994).

This study adds to the body of literature that demonstrated a negative effect of exposure to infectious diseases in early childhood on old-age mortality (Bengtsson and Lindstrom, 2000; Blackwell et al., 2001; Costa, 2003; Doblhammer and Vaupel, 2001; Elo and Preston, 1992; Finch and Crim- mins, 2004). We study the effect of exposure to waterborne diseases in early life on old-age mortality. The introduction of water treatment reduced exposure to infectious diseases for children. We aim to determne if adopting these systems which plausibly lowered exposure water-born diseases in early childhood in American cities at the beginning of the 20th century resulted in reduced mortality at old age in the second half of the 20th century. The gradual role out of these systems meant (1) at a point in time, children in different cities had different access to clean

[1] Other public health interventions including modern sewerage systems and public refuse management contributed as well.

drinking water; and (2) within a city, the infectious disease burden was lowered when children were different ages.

Costa (2000, 2003) demonstrated that growing up in a big city significantly increased the risk of death later in life from chronic diseases among Union Army veterans, in particular from heart disease. Costa (2003) also studied the effect of hospitalization due to typhoid fever while in the army on old-age mortality and the effect was not a significant predictor of death in general, nor from any specific cause. For these two studies, diseases at early in life was not isolated from other negative factors such as of growing up in a big city, i.e. worse housing conditions and worse diet composition as compared to the countryside, or from other factors related to typhoid in the veterans' sample. Water purification systems were introduced in large American cities and when they were introduced relative to a birth cohort seems plausibly exogenous. Troesken (2004) establishes that cohorts born in large American cities just after water treatment was widely implemented would had lower mortality and that this was linked to a reduction in water borne disease deaths. Following on Troesken (2004) we ask did exposure to water treatment lower mortality at old age? That is was there an improvement in long term health from these public health innovations?

Data, Data Limitations and Methods

The data consists of Medicare Part B enrollment data matched by social security number to records from The Numerical Identification Files (NUMIDENT) from Social Security Administration. Eligibility to enroll in the Medicare Part B program is based on entitlement to Medicare Part A; for a majority of Americans when they turn age 65. Often demographers use Vital Statistics data to measure the number of deaths at any age and rely on Census samples to measure the population at risk. A major advantage of the Medicare-NUMIDENT data is that estimates of mortality can be

[2] This is known as the *Mills-Reincke phenomenon*,

formed from a single data source – individuals enroll at age 65 and are followed until they die. In addition, age reporting is more accurate than in census or death certificated because proof of age is required to enroll for Medicare benefits (Kestenbaum, 1992). The data provides nearly complete coverage of the US population: According to Elo et al. (2004) in 2001 close to 93% of elderly were enrolled in the Medicare Part B program. There are over 70 million records in the data-set. Medicare data provides basic demographic data on race, sex, and age an age at death. We focus on the 2 million records for individuals born between 1900 and 1915.

The NUMIDENT is a computerized abstract of SS-5 form, which is the form people fill out when applying for a social security number. The Numident file has the person's exact date of birth, race, and a 12 character string that identifies the individual's town of birth and a two character string that identifies the state of birth if applicable. We developed an algorithm that matches these objects to place names recorded in the U.S. Geological Service's Geographic Names Information System (GNIS). The GNIS is the master list of all populated place names in the U.S. both current and historic, and includes geographic features including the longitude and latitude of each place. The list of places was obtained from http://geonames.usgs.gov/domestic/download_data.htm. Our goal was to accurately assign each 12-letter string to its most likely match among the populated places and counties in the person's birth state based on the percent of letters that are common between the strings. Each unique 12-letter string that has 10 or more occurrences in the data was hand-checked for accuracy. Approximately 15% of men and 23% of women had missing information on birthplace, similar to numbers reported by Dupre et al. (2012); Elo and Preston (1997); Lauderdale and Kestenbaum (2002). Applying the same data source to study differences in survival among native-born and foreign-born populations, Dupre et al. (2012) reported no significant bias due to missing information on place of birth.

An important limitation study is that the coverage rates for the relevant years of this study are very low and likely very select. Some detail on data construction is in order. While the Medicare program started already in 1966 and the NUMIDENT in 1937, the computerization of the NUMIDENT occurred only in 1976. At this time, the Social Security Administration held SS-5 cards for individuals who had not filed for benefits in Washington D.C. However, once an individual filed for benefits the SS-5 card was sent to the individual's regional social security office where administration of the program was handled. This means that in 1975 when the NUMIDENT was computerized only cards residing in Washington D.C. were data entered. For example, for individuals born in 1900, only individuals who were alive and had not filed for benefits (now age 75) were computerized; for individuals born in 1901, only those ages 74 who had not filield for benefits were computerized. Since individuals normally file for benefits at ages 62 or 65, the coverage rates for these cohorts are very low. In addition, the older a cohort the more select the cohort. Which way the selection goes is not clear but we offer some empirical evidence on this below.

Table 1:
Estimated Coverage Rates in the Medicare-SSA data

Birth Year	Men	Women
1900	3.4%	18.5%
1901	3.1%	15.8%
1902	4.1%	21.7%
1903	3.9%	19.1%
1904	4.2%	19.9%
1905	5.0%	18.4%
1906	5.4%	19.0%
1907	6.1%	19.9%
1908	7.7%	19.7%
1909	7.7%	19.5%
1910	9.2%	20.1%
1911	15.2%	24.1%
1912	37.7%	35.3%
1913	40.4%	38.2%
1914	48.2%	47.8%
1915	72.4%	68.8%
1916	74.2%	73.4%
1917	78.4%	77.1%
1918	78.2%	70.7%
1919	81.5%	73.4%
1920	78.0%	77.7%

Table 1 shows the coverage rates. To calculate these rates we count the number of men and women aged 75 in our data to form our numerator and estimate the number of men and women aged 75 in the 1976-1996 CPS to form our denominator. These estimates are presented in Table 1. After 1915 the coverage rates are generally 70%-80% (we would expect the maximum coverage rate to be approximately 93% as this is the fraction covered by Medicare part B). But clearly the coverage rate is extraordinarily low for the early cohorts, especially for men. This reflects the rare occurrence that men reached age 75 without filing for Social Security under their SSA number. More women remained in the file at age 75 reflecting that many women did not qualify for Social Security under their own SSA number in these cohorts.

Even with what are obviously large coverage issues and large potential selection issues we believe our estimates still have internal validity. For any birth cohort we are always comparing within-city old age mortality for those that were born in after the adoption of water treatment to those who were born before. However, it is important to recognize that these treatment effects are estimated on select populations whose selection is changing over birth cohorts. In addition, it means that only limited statistics can be estimated across birth cohorts. For all birth cohorts we can estimate the probability of living t years given an individual was alive at age 75. But statistics at earlier ages can only be estimated for later cohorts. Because of the exceptionally poor coverage rates, as well as the higher fraction of women living to age 75 in these cohorts, we emphasize our analysis for women (but we calculate statistics for men as well).

Early exposure to water-born infectious diseases in the study is approximated by date and place of birth in relation to the date a water treatment system was installed on a massive scale in the city of birth. Information on the exact year when mechanical water filtration was introduced in 36 US cities and when water chlorination started in 41 cities between 1890 and 1915 was derived from Department of Commerce, Bureau of Census (1916). The five methods of water treatment listed in the reference are: sedimentation, coagulation, slow sand filtration, mechanical filtration and chemical sterilization. We limit our study to the effect of mechanical filtration and chemical sterilization, as other methods proved not to be fully efficient in removal of bacteria and viruses from contaminated water: According to the study of Gimbel and Clasen (1998), cited after World Health Organization (2004), coagulation and sedimentation together removed from 27 to 74% viruses and 32-87% bacteria. The quoted numbers were collected in water-treatment plants at the end of the 20th century and one can expect that efficiency of the methods at the beginning of the century was much lower than that. Appendix A-1 reports the table from the 1916 Census report.

We estimate a series of linear probability models. We concentrate on one statistic that can be calculated for all birth cohorts born between 1900 and 1920 – the probability of living to age 80 given an individual has lived to age 75. For cohorts born between 1900 and 1920, 86% of women and 90% of men who lived to age 75 will die before their 80th birthday. We restrict our sample to the set of cities that adopt either chemical sterilization of mechanical water treatment by 1914. In order to rely on within city variation to identify the effect of water treatment we include city fixed effects in the model. We also only use observations within an 8-year window surrounding the adoption of water treatment. That is the pre-period consists of the four years prior to water treatment and the post period is being born in the year of water treatment or the 3 years following it. We control for birth cohort either with birth cohort fixed effects or with a quadratic in year of birth. In order to take into account the coverage issues, we weight each observation by the inverse probability of inclusion in the sample.

Results

Table 2 displays the main results in the paper. We first estimate the model for white women including city of birth fixed effects (Column 1). Over this period 14% of women aged 75 survived to age 80. Our estimates suggest that the advent of water treatment lowered this mortality rate by 1.39 percentage points. This is 10% increase in the probability of survival at old ages. Column 2 shows that modeling the change in cohort mortality as a quadratic in birth year give results that are very close to the fixed effect results. It is clear in this specification that survival at older ages is clearly growing with birth cohort at the rate of over 1 percentage point a year for women. This large growth in cohort survival reflects a real growth in survival but also a change in the selectivity of women in our sample as the coverage rates grow over time. Given the strong pattern of cohort growth in survival we suspect that selection is playing a large role. Column 3 shows the results for men. Given the large coverage issues (and small sample sizes) it is of note that our estimated effects for men are comparable to those for women and just fail to reach statistical significance.

The last column simply pools men and women to try and estimate the impact of water treatment with more precision. Given that the point estimates for the two groups are so close it is not surprising that the estimate combining men and women is also similar.

Table 2:
Estimated Effects of Water Treatment of 5-Year Survival

VARIABLES	(1) White Women YOB FE Pr(Death by 80)	(2) White Women Pr(Death by 80)	(4) White Men Pr(Death by 80)	(3) All Whites Pr(Death by 80)
Birth Year -1900		0.0132***	0.00586***	0.00605**
		(4.887)	(2.929)	(2.910)
(Birth Year -1900)^2		0.00136***	0.00121***	0.00121***
		(9.688)	(5.675)	(7.100)
Born After Water Treatment	**0.0139****	**0.0114***	**0.0119**	**0.0117*****
	(2.124)	**(2.096)**	**(1.639)**	**(3.546)**
Female				-0.00500**
				(-2.259)
Female*(Birth Year -1900)				0.00695***
				(4.981)
Female*(Birth Year -1900)^2				0.000145
				(1.622)
Constant	0.000***	-0.003***	0.024***	0.011***
	(77.99)	(73.37)	(74.97)	(106.7)
Observations	55,173	55,173	19,416	74,589
R-squared	0.165	0.165	0.120	0.153
Mean of Dependent Variable	0.1434	0.1434	0.0985	0.1231

Robust t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Summary and Discussion

The research question of the study concerned long-term effect of early exposure to water-borne diseases and was addressed in a set of four research hypotheses. The hypotheses tested the overall effect of access to a municipal clean water source in early childhood on old-age mortality. Early- life exposure to water-born infectious diseases was approximated by year of birth in relation to the date a water treatment system was installed on a massive scale in the city of birth. Two methods of drinking water treatment introduced at the beginning of the 20th century in large American cities were taken into account in the study: mechanical filtration and chemical sterilization. We find that these public health interventions at the beginning of the 20th century may have had lasting effects on human health. While the work here is preliminary, we believe that this is among the first attempts to measure the long-term impact of these important public health interventions in the early 20th century. If these results hold up, it suggests important benefits to public health interventions that would be difficult to take into account when considering whether such interventions meet cost-benefit criteria. There is potentially an important lesson – effects on human health early in the life course likely raise health throughout life. This increases the likelihood that various public health interventions are cost effective.

References

Bengtsson, T. and M. Lindstrom (2000). Childhood misery and disease in later life: The effects on mortality in old age of hazards experienced in early life, Southern Sweden, 1760-1894. *Population Studies 54*(3), 263–277.

Blackwell, D. L., M. D. Haywardb, and E. M. Crimmins (2001). Does childhood health affect chronic morbidity in later life? *Social Science and Medicine 52*, 1269–1284.

Caygill, C. P. J., M. J. Hill, M. Braddick, and J. C. M. Sharp (1994). Cancer mortality in chronic typhoid and paratyphoid carriers. *The Lancet 343*, 83–84.

Clay, K. and W. Troesken (2006). Deprivation and disease in early twentieth century America. *NBR Working Papers 12111*.

Coale, A. J. and E. E. Kisker (1986). Mortality crossovers: Reality or bad data? *Population Studies 40*(3), 389–401.

Costa, D. L. (2000). Understanding the twentieth-century decline in chronic conditions among older men. *Demography 37*, 53–72.

Costa, D. L. (2003). Understanding mid-life and older age mortality declines: evidence from Union Army veterans. *Journal of Econometrics 112*, 175–192.

Cutler, D., A. Deaton, and A. Lleras-Muney (2006). The determinants of mortality. *Journal of Economic Perspectives 20*, 97–120.

Cutler, D. and M. Grant (2005). The role of public health improvements in health advances: The twentieth-century United States. *Demography 42*, 1–22.

Cutler, D. M., E. L. Glaeser, and J. L. Vigdor (1999). The rise and decline of the american ghetto. *Journal of Political Economy 107*(3).

Department of Commerce, Bureau of Census (1916). *General Statistics of Cities: 1915*. Washington: Government Printing Officer.

Doblhammer, G. and J. W. Vaupel (2001). Lifespan depends on month of birth. *Proceedings of the National Academy of Sciences 98*, 2934–2939.

Dupre, M. E., D. Gu, and J. W. Vaupel (2012). Survival differences among native-born and foreign-born older adults in the united states. *PloS one 7*(5), e37177.

Elo, I. T. and S. H. Preston (1992). Effects of early-life conditions on adult mortality: a review. *Population Index 58*, 186–212.

Elo, I. T. and S. H. Preston (1997). Racial and ethnic differences in mortality at older ages. *Racial and ethnic differences in the health of older Americans*, 10–42.

Elo, I. T., C. M. Turra, B. Kestenbaum, and B. R. Ferguson (2004). Mortality among elderly hispanics in the United States: Past evidence and new results. *Demography 41*, 109–128.

Ferrie, J. P. and W. Troesken (2008). Water and Chicago's mortality transition, 1850-1925. *Explorations in Economic History 45*, 1–16.

Finch, C. E. and E. M. Crimmins (2004). Inflammatory exposure and historical changes in human life-spans. *Science 305*, 1736–1739.

Gimbel, R. and J. Clasen (1998). International report. *Water Supply-International Water Supply Association 16*(1), 203–207.

Kestenbaum, B. (1992). A description of the extreme aged population based on improved medicare enrollment data. *Demography 29*, 565–580.

Kestenbaum, B. (2000). Commentary on 'a methodology for identyfying married couples in medicare data'. *Demography 37*, 247–249.

Lauderdale, D. S. and B. Kestenbaum (2002). Mortality rates of elderly asian american populations based on medicare and social security data. *Demography 39*, 529–540.

Manton, K. G., C. H. Patrick, and K. W. Johnson (1987). Health differentials between blacks and whites: recent trends in mortality and morbidity. *The Milbank Quarterly*, 129–199.

Preston, S. H., I. T. Elo, I. Rosenwaike, and M. Hill (1996). African-american mortality at older ages: Results of a matching study. *Demography 33*(2), 193–209.

Sedgwick, W. T. and J. S. MacNutt (1910). On the mills-reincke phenomenon and hazen's theorem concerning the decrease in mortality from diseases other than typhoid fever following the purification of public water-supplies. *Journal of Infectious Diseases 7*(4), 489–564.

Social Security Administration (2006). Medicare Part B and NUMIDENT data at Duke University.

Troesken, W. (2004). *Water, Race, and Disease*. Cambridge, Massachusetts: MIT Press.

U.S. Department of Commerce,Census Bureau (1994). *Geographic Areas Reference Manual*. Washington, DC, U.S. Government Printing Office.

Vaupel, J. W. and A. I. Yashin (2006). Unobserved population heterogeneity. In G. Caselli, J. Vallin, and G. Wunsch (Eds.), *Demography. Analysis and Synthesis. A Treatise in Population Studies*, Volume 1, pp. 271–278. Elsevier Academic Press.

World Health Organization (2004). *Water Treatment and Pathogen Control: Process Efficiency in Achieving. Safe Drinking Water. Edited by Mark W LeChevallier and Kwok-Keung Au.* Published by IWA Publishing, London, UK.

DIFFERENT PURIFICATION PROCESSES NOW IN USE.

YEAR[1] INSTALLED.	Sedimentation.	Coagulation.	Slow sand filtration.	Mechanical filtration.	Chemical sterilization.
1853	Washington, D. C.				
1879	Louisville, Ky.				
1883	Council Bluffs, Iowa				
1889	Dallas, Tex				
1890	Omaha, Nebr	Omaha, Nebr		Oshkosh, Wis	
1890	Oshkosh, Wis				
1892	Atlanta, Ga				
1893			Lawrence, Mass		
1894	Knoxville, Tenn	Knoxville, Tenn	Altoona, Pa		
1894			Knoxville, Tenn		
1896		Cedar Rapids, Iowa		Charlotte, N. C.	
1896				Cedar Rapids, Iowa	
1898	Macon, Ga		Augusta, Ga		
1898			Macon, Ga		
1899	Albany, N. Y		Albany, N. Y	Norfolk, Va	
1900	Kansas City, Mo	Kansas City, Mo			Mobile, Ala.
1902	Philadelphia, Pa		Philadelphia, Pa	Philadelphia, Pa	
1902				Binghamton, N. Y	
1902			Providence, R. I		
1902			Austin, Tex		
1903			Washington, D. C	New York, N. Y	
1903			Reading, Pa		
1903			Yonkers, N. Y		
1904	St. Louis, Mo	St. Louis, Mo		Knoxville, Tenn	
1904		Atlanta, Ga			
1905	Youngstown, Ohio	Harrisburg, Pa		Youngstown, Ohio	Harrisburg, Pa.
1905	Wilmington, Del	Charlotte, N. C		Harrisburg, Pa	
1905	Harrisburg, Pa	Columbia, S. C		Columbia, S. C	
1905	Columbia, S. C				
1906			New York, N. Y		
1907	Oklahoma City, Okla	Lorain, Ohio		San Diego, Cal	
1907				Lorain, Ohio	
1908	Pittsburgh, Pa	Cincinnati, Ohio	Pittsburgh, Pa	Cincinnati, Ohio	Jersey City, N. J.
1908	Cincinnati, Ohio	New Orleans, La	Wilmington, Del	New Orleans, La	Columbus, Ohio.
1908	New Orleans, La			Columbus, Ohio	Omaha, Nebr.
1908	Nashville, Tenn			McKeesport, Pa	Charlotte, N. C.
1908	McKeesport, Pa				
1909	Richmond, Va	Louisville, Ky	Kansas City, Kans	Louisville, Ky	Nashville, Tenn.
1909	Kansas City, Kans	Richmond, Va		Albany, N. Y	Kansas City, Kans.
1909		Nashville, Tenn		Kansas City, Kans	
1909		Kansas City, Kans			
1910	Springfield, Mass	Washington, D. C	Springfield, Mass	Toledo, Ohio	Pittsburgh, Pa.
1910	Pueblo, Colo	Springfield, Mass		Atlanta, Ga	Milwaukee, Wis.
1910		Pueblo, Colo			Lima, Ohio.
1910		Council Bluffs, Iowa			Council Bluffs, Iowa.
1911		Niagara Falls, N. Y		Montgomery, Ala	Cincinnati, Ohio.
1911				Niagara Falls, N. Y	Kansas City, Mo.
1911					Trenton, N. J.
1911					Albany, N. Y.
1911					Lincoln, Nebr.
1911					Muskogee, Okla.
1911					Niagara Falls, N. Y.
1912	Flint, Mich	Evansville, Ind		Grand Rapids, Mich	New York, N. Y.
1912	Newport, Ky			Fort Worth, Tex	Chicago, Ill.
1912				Evansville, Ind	St. Louis, Mo.
1912				Flint, Mich	Detroit, Mich.
1912					Tacoma, Wash.
1912					Wilmington, Del.
1912					Evansville, Ind.
1912					Flint, Mich.
1912					Cedar Rapids, Iowa.
1913	Minneapolis, Minn			Minneapolis, Minn	Philadelphia, Pa.
1913					Cleveland, Ohio.
1913					Louisville, Ky.
1913					Hartford, Conn.
1913					Duluth, Minn.
1913					Portland, Me.
1913					Auburn, N. Y.
1913					Lynchburg, Va.
1914	Trenton, N. J	Dallas, Tex		Baltimore, Md	Buffalo, N. Y.
1914	Decatur, Ill	Trenton, N. J		Dallas, Tex	Dallas, Tex.
1914		Albany, N. Y		Trenton, N. J	Bay City, Mich.
1914		Decatur, Ill		Erie, Pa	Decatur, Ill.
1914				Decatur, Ill	Columbia, S. C.
1914				Waco, Tex	
1915				St. Louis, Mo	

[1] Calendar year.