Aggregating Fetal Origins

James Feyrer
Dartmouth College

July 23, 2024

Abstract

Over the last 70 years the world has seen a decrease in infant mortality and a rise in life expectancy. There is a substantial body of research that shows that health conditions in utero and early childhood have a significant impact on adult health. This paper quantifies this connection using yearly cross country data on age specific mortality rates. A 10% movement in the age 0-1 mortality rate faced by a cohort results in a 2-3% higher mortality rates for that cohort every year for the rest of their lives. Due to the rich set of controls, the identifying variation is relatively small year to year variations in health conditions faced by specific cohorts born that year. Replacing the overall mortality results with mortality shock specifically from respiratory disease yields similar results. Aggregating the main result, infant mortality increases can potentially account for over 60% of the country level increases in life expectancy at age 10 over the last 70 years.

Introduction

Over the last 20 years a large and growing literature in economics has focused on the long run impacts of early life health conditions. Almond and Currie (2011) and Almond, Currie and Duque (2018) provide comprehensive accounts of this work. This work goes back to Barker who has been a leading proponent of the "fetal origins hypothesis," which suggests that the conditions faced in utero have implications for adult health. Barker has published a series of papers connecting low birth weight to later heart disease, diabetes and other adult conditions.[1] There is also a substantial demographics literature that models the

[1]Barker and Osmond (1986), Barker (1990), Barker (1995)

impact of infant health on mortality. Lleras-Muney and Moreau (2022) is a good recent example that contains a thorough review of this literature from both the demography and economics perspective.

The volume of papers and the quality of results backing the fetal origins hypothesis suggest that we should be able to observe this in the aggregate data. Changes that we have observed in infant health and survival over the last 70 years should be having a significant impact on the health of the adult populations over time. How much do better health conditions in the year of your birth reduce your mortality risk throughout the rest of your life? How much of improving health worldwide can therefore be attributed to health conditions decades in the past? In contrast to previous literature, this paper will take a reduced form macro approach.

I focus on age specific mortality rates compiled by the UN for each country in the world between 1950 and 2019.[2] This revision of the World Population Prospects is the first to use yearly data. The richness of this data allows me to track the mortality profile of cohorts at the country level over time and connect them to health conditions in the year of their birth. The first part of the paper regresses age specific mortality rates by country by year against the country specific mortality rate faced by each cohort in their first year of life. This is not a perfect measure of fetal conditions but should provide a very useful proxy.

Age by year fixed effects control for age specific mortality improvements worldwide. Country by year fixed effects control for country level shocks to health that impact all age groups. Country by age fixed effects can control for differences in the age profile of mortality by country. The age profile of mortality at the country level may change over time. I would ideally control for country by age by year, but this would obviously uniquely identify each observation. I include country by age by 10 year groups to deal with longer run trends in the age profile by country.[3] This does not change the results significantly

[2](United Nations 2022)

[3]I can reduce the year groupings to 5 or even 2 years and it does not significantly change the results.

compared to simple country by age controls.

The identifying variation that remains is from comparing the age 0-1 mortality rate at birth for a cohort to adjacent cohorts in the same country while accounting for yearly movements in overall mortality within that country and age specific mortality changes happening worldwide. Essentially I am looking at idiosyncratic country level shocks to age 0-1 mortality faced by individual cohorts.

I find an elasticity of roughly 0.27 from age 0-1 mortality to lifetime yearly risk of death. A 10% decrease in age 0-1 mortality for your cohort is associated with a 2.7% reduction in mortality every year after. The result is specific to the age 0-1 mortality rates of a cohort in the year of birth. The age 0-1 mortality rate of surrounding cohorts does not have a significant impact. Cohort mortality at ages 1-10 have an impact on later mortality that is an order of magnitude smaller than cohort mortality at age 0-1. The results suggest that the age 0-1 mortality rate is crucial for later mortality risk.

The majority of identifying variation is from small mortality movements of +/- 5% that are consistent with random fluctuation in respiratory disease rates. To verify this I replace the overall age 0-1 cohort mortality with mortality due specifically to respiratory disease. Though this reduces the sample size in both the country and time dimension, I find a similar impact on later life mortality to the base results.

The second part of the paper examines counterfactual scenarios where I start with the 1950 age mortality profile by country and ask what would happen to life expectancy over time if reductions in age 0-1 mortality changed the lifetime mortality profile in the way suggested by the empirical estimates. Assuming that my estimates can be applied to longer run trends (an admittedly large assumption!), These scenarios suggest that well over half of the life expectancy reductions experienced over the last 70 years are the result of cohorts that have been healthier since the year of their birth.

1 Data

The data are from the UN World Population Prospects 2022 United Nations (2022). This is the first year that the WPP revision has included yearly data and this led to significant revisions to their historical data. The entirety of the analysis falls on age specific mortality by country by year from 1950 to 2019.[4] I am relying on the World Bank's calculations of age specific mortality by year.[5] I will also use population by age by year as measured on July 1.

It is worth noting that the mortality data is variable in quality. The highest quality estimates are from countries with a complete set of vital statistics (i.e. 100% of individual births and deaths are recorded). The World Bank distinguishes between "empirical" and "model based" estimates of mortality. While I will report results from both type of mortality estimates, the empirical estimates will be used in the main specifications.

The full dataset includes 235 locations with both the model and empirical estimates. The documentation for the WPP includes information on which country's mortality data is based on vital records, what year they started, and their completeness. The subsample HQ1960 is any country with complete vital records going back to 1960 – 125 countries.

This includes some very small islands like the Falklands which has less than 5,000 people. At cohort sizes of less than 1,000 people death rates can be low enough that 1-2 deaths a year can significantly move age specific death rates. The final subsample eliminates any location with an average population over the sample of less than 100,000. This final sample includes 98 countries.

Each of these succeeding samples should have smaller measurement error. Model based estimates will be less specific to particular dates and places (i.e. it is hard to know that

[4]The data are available through 2021, but I exclude these due to the COVID-19 epidemic. This has no significant impact on the results.

[5]Specifically, the mx variable from the Life Tables data found here `https://population.un.org/wpp/Download/Standard/CSV/`

age 0-1 mortality is high in a particular year in a particular country without specific data). Low population countries will similarly have more year to year variation in mortality rates due to very low numbers of recorded deaths in some age groups. These data will naturally have more measurement error and will tend to bias the estimates towards zero.

In addition to the World Bank mortality measures I use age specific death rates due to respiratory infections and tuberculosis from the Institute for Health Metrics and Evaluation (IHME) from 1990-2009.[6]

2 Results

The basic estimating equation for all regression is

$$ln(MORT_{loc,age,year}) = \beta \times ln(MORT_{loc,0,year-age}) + \gamma_{loc,year} + \gamma_{age,year} + \gamma_{loc,age,decade} + \epsilon \quad (1)$$

The left hand side variable is the mortality rate by age by country by year. The right hand side variable is the mortality rate at ages 0-1 experienced by the cohort whose death rate is on the left hand side. For example, if the data point is the mortality rate for 34 year olds in the US in 2004, the right hand side will be the age 0-1 mortality rate in 1970, the year of their birth. For the 38 year old as measured in 1984, the right hand side will be the age 0-1 mortality rate in 1967. All data are in logs, so the results represent percentage shifts in death rates.

All regressions have a full set of age X year, country X year, and country X age X decade fixed effects.[7] Country by year effects isolate overall changes in the health environment in a country over time on a yearly basis. Age by year effects account for common global shocks to mortality and they are age specific. These controls are equivalent to controlling

[6]Downloaded from `https://vizhub.healthdata.org/gbd-results/`.

[7]The country X age X decade fixed effects can be replace by a two year window without significantly impacting the results.

for worldwide birth cohorts. Finally country by age controls allow for a different age profile of mortality across countries.

Identification is therefore coming from changes in age 0-1 mortality rates for a cohort relative to adjacent cohorts from the same country, accounting for overall health improvement within the country and around the world.

Table 1 shows the impact of age 0-1 mortality on the profile of mortality of a cohort over the rest of their observed lives. The results suggest that for a 10% increase in cohort age 0-1 mortality, the cohort experiences a 2.7% increase in their yearly mortality rate on average for the rest of their lives.

Moving across the columns to smaller and lower measurement error samples significantly increases the coefficient, exactly what we would expect. Columns (4-6) include the size of the cohort as an additional regressor. This shows that the results are being driven by variation in the number of deaths and not by cohort size.

Figure 1 shows a visualization of the results from Table 1, Column (3). Mortality rates and cohort age 0-1 mortality rates have been demeaned by the same set of controls used in the regressions. This figure shows that identification is not being driven by outliers. Over 90% of the residual changes to age 0-1 mortality are smaller than 10%. The results are very stable if you trim the tails of the shock distribution.

Table 2 shows that the results are specific to the cohort mortality rate at age 0-1. Column (1) adds the cohort mortality rate for each year from birth to age 5. This shows that the impact of higher childhood mortality is highly concentrated on year 0. Column (2) includes the age 0-1 mortality rates for surrounding cohorts. Again, the results are being driven by idiosyncratic shocks experienced by a particular cohort around the time of their birth.

The results so far have assumed that the mortality shift is uniform across ages. It is possible to estimate the age specific shift in mortality due to age 0-1 mortality shocks with

a full set of age interactions.

$$MORT_{loc,age,year} = \sum_{age=1}^{70} \beta_{age} \times MORT_{loc,0,year-age} + \gamma_{loc,year} + \gamma_{age,year} + \gamma_{loc,age,decade} + \epsilon \quad (2)$$

Figure 2 shows the profile of mortality shifts across the lifetime with a 95% confidence interval. The impact is somewhat larger at ages 1-10 and settles into a constant around 0.2 for the rest of the age profile.

Table 1: The impact of age 0-1 mortality on later cohort mortality

	(1) Full	(2) HQ1960	(3) HQ1960p200	(4) Full	(5) HQ1960	(6) HQ1960p200
ln(cohort mx 0-1)	0.145	0.149	0.267	0.145	0.149	0.267
	(0.0224)	(0.0276)	(0.0402)	(0.0224)	(0.0275)	(0.0402)
ln(cohort pop 0-1)				-0.00237	-0.00867	0.00151
				(0.00899)	(0.0106)	(0.0193)
R2	0.997	0.996	0.996	0.997	0.996	0.996
Within R2	0.0208	0.0222	0.0381	0.0208	0.0222	0.0381
Observations	568052	300875	235886	568052	300875	235886
Countries	236	125	98	236	125	98
Years	68	68	68	68	68	68
Cohorts	69	69	69	69	69	69

Standard errors in parentheses
Regressions include ageXyear, locationXyear, and locationXyearX10yeargroup

Full – all available data
HQ1960 – based on complete vital records as of 1960
HQ1960p200 – complete vital records + average population > 200k

Table 2: The impact is focused on Age 0-1

	(1) ln(mortality)		(2) ln(mortality)	
ln(cohort mx 0-1)	0.171	(0.0336)	0.303	(0.0800)
ln(cohort mx 1-2)	0.00888	(0.0120)		
ln(cohort mx 2-3)	0.0137	(0.00620)		
ln(cohort mx 3-4)	0.0140	(0.00731)		
ln(cohort mx 4-5)	0.0163	(0.00465)		
ln(cohort mx 5-6)	0.0171	(0.00701)		
L3.ln(cohort mx 0-1)			0.00696	(0.0113)
L2.ln(cohort mx 0-1)			-0.00480	(0.0164)
L.ln(cohort mx 0-1)			-0.0379	(0.0225)
F.ln(cohort mx 0-1)			-0.0124	(0.0237)
F2.ln(cohort mx 0-1)			0.00875	(0.0189)
F3.ln(cohort mx 0-1)			0.0224	(0.0198)
R2	0.996		0.996	
Within R2	0.0298		0.0370	
Observations	196784		209426	
Countries	98		98	
Years	62		64	
Cohorts	63		65	

Standard errors in parentheses
Regressions include ageXyear, locationXyear, and locationXyearX10yeargroup

Figure 1: Binscatter of death rates against cohort age 0-1 mortality

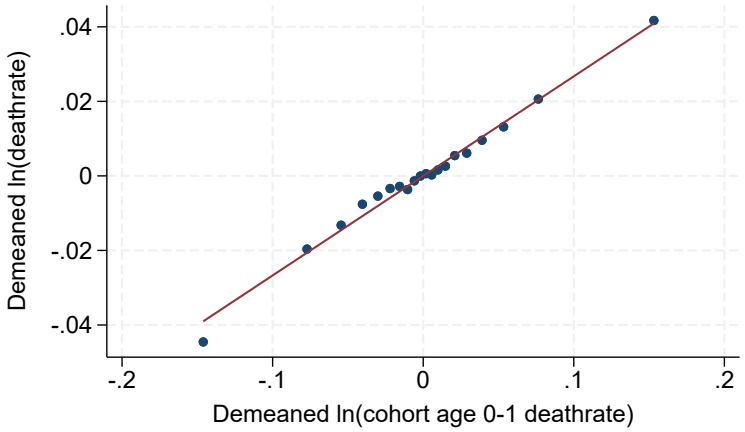
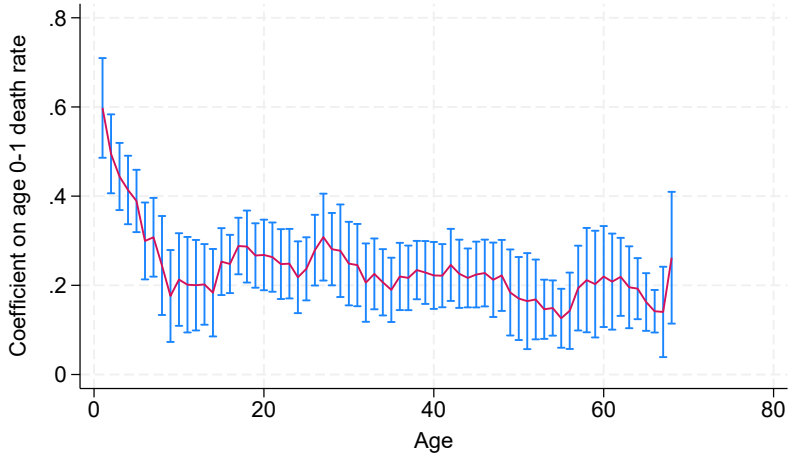


Figure 2: The impact of age 0-1 mortality on cohort mortality by age



3 Respiratory disease as a random health shock

What is driving the identification in the results we have seen so far? Figure 1 shows that the majority of the residual variation in age 0-1 mortality rates is relatively small, with most shocks being +/- 10% or smaller. This figure also shows that the slope if relatively constant over the entire range of shocks. What sort of events would drive shocks in this range? One possibility is yearly variation in seasonal respiratory disease.

Table 3 shows the results from replacing overall age 0-1 mortality rates on the right hand side with age 0-1 mortality rates from respiratory disease from the IHME. These data only run from 1990 - 2019 so I cannot examine the impact past the age of 30. This means that the sample size is significantly smaller. The data are also noisier than the overall death data because cause of death is less certain than the fact that a death occurred.[8] Despite these challenges, the results are quite clear. The noisier disease measure is still a significant predictor of later mortality, with an elasticity a bit more than 1/2 that of the overall mortality measure. Figure 3 is a binscatter visualizing the results from Table 3.

It is also instructive to look at the size of shocks across countries. The fixed effects ensure that the shocks are roughly mean zero within a country so the interesting question is what countries have a higher deviation of shocks over time? This may provide some insight into the sources of identifying variation. I averaged the demeaned age 0-1 mortality rate by cohort and took the standard deviation across cohorts within a country.

Figure 4 shows the scatter of absolute latitude and the standard deviation of the identifying variation at the country level from the main results. Being further from the equator is associated with larger cohort to cohort variation in age 0-1 death rates. Note that the outliers are of two types. First, countries which experienced significant conflict over the sample period like North Korea and Cambodia and second, low population nations like Iceland, Malta, and Luxembourg.

[8] According to the IHME, the 95% confidence interval on these data is +/- 27%.

Table 3: The Impact of Cohort Age 0-1 Respiratory Disease Deaths

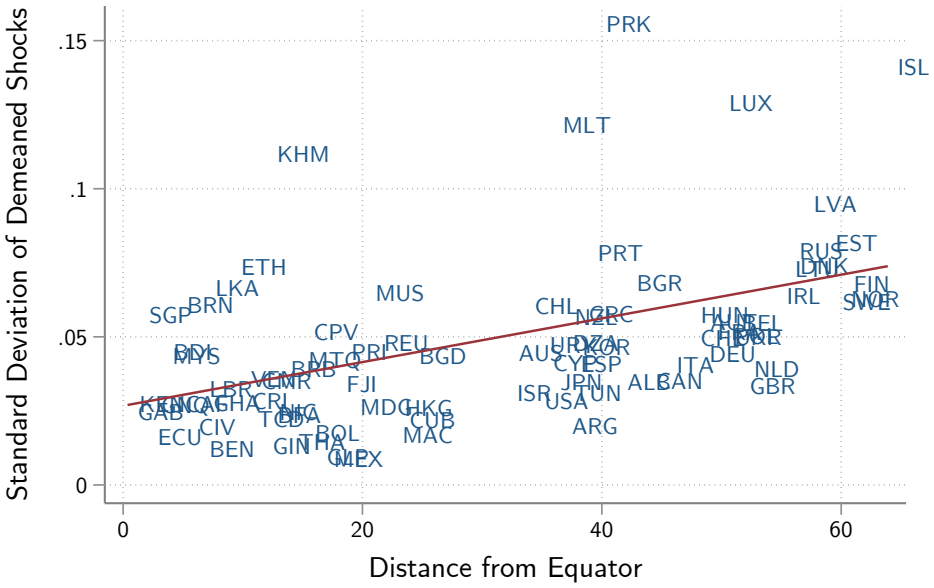
	(1) ln_mx	(2) ln_mx	(3) ln_mx	(4) ln_mx
ln(cohort mx 0-1)	0.418		0.416	
	(0.0889)		(0.0893)	
ln(Age 0-1 Resperatory Disease Death Rate)		0.167		0.164
		(0.0325)		(0.0336)
ln(cohort pop 0-1)			-0.0469	-0.0513
			(0.0653)	(0.0601)
R2	0.991	0.991	0.991	0.991
Within R2	0.0429	0.00879	0.0431	0.00898
Observations	38359	38359	38359	38359
Countries	89	89	89	89
Years	28	28	28	28
Cohorts	29	29	29	29

Standard errors in parentheses
Regressions include ageXyear, locationXyear, and locationXyearX10yeargroup

Figure 3: Binscatter of Cohort Mortality versus Cohort Age 0-1 Respiratory Disease Death



Figure 4: The Variability of shocks versus latitude



This result is consistent with the idea that much of the identifying variation is coming from random year to year fluctuations in disease. There is an extensive literature that shows that colder, higher latitude countries have higher incidence and variation in seasonal respiratory disease.[9]

[9]see Cannell, Vieth, Umhau, Holick, Grant, Madronich, Garland and Giovannucci (2006) and Bloom-Feshbach, Alonso, Charu, Tamerius, Simonsen, Miller and Viboud (2013) on the relationship between the seasonality of RSV and influenza by latitude.

4 Aggregate Impacts on Life Expectancy

How much of life expectancy change over the last 70 years can be accounted for by improvements in age 0-1 mortality that occurred when the current population was born? To answer this question I start with country mortality age profiles in 1950 and predict how age specific mortality would change under different assumptions for the path of age 0-1 mortality. I can then compute life expectancy by age and compare this to the actual life expectancies we observe in the data.

The first approach is to start with the 1950 Mortality profile and predict 2019 age specific mortality using the age 0-1 mortality rates of cohorts alive in 2019 and estimates on the impact of age 0-1 mortality on lifetime mortality from the previous section. This is equivalent to a prediction from equation (1) stripping out all of the fixed effects. The only inputs to the prediction are the initial age mortality profile and the change in age 0-1 mortality from 1950 to 2019. For the following equations, $M_{loc,age,year}$ is the mortality rate by country, by age, by year.

$$ln(M_predict_{loc,age,2019}) = ln(M_{loc,age,1950}) + \beta_{age} \times (ln(M_{loc,0,2019-age}) - ln(M_{loc,0,1950})) \quad (3)$$

The second scenario calculates what age specific mortality would be in 2019 if all the cohorts alive today had experienced the age 0-1 mortality faced by the 2019 cohort. This can be seen as a prediction for where age specific mortality would settle in the long run if age 0-1 mortality stayed constant at the 2019 level.

$$ln(M_predict_{loc,age,2019}) = ln(M_{loc,age,1950}) + \beta_{age} \times (ln(M_{loc,0,2019}) - ln(M_{loc,0,1950})) \quad (4)$$

The final thought experiment asks what age specific mortality would be in 2019 if we started with country specific 1950 mortality by age profiles and reduced age 0-1 mortality

to the lowest levels observed in the world in 2019 (roughly 0.2%) in all countries. This can be seen as a prediction for age specific mortality if all countries moved to the current frontier of age 0-1 mortality.

$$ln(M_loc, age, 2019) = ln(M_{loc,age,1950}) + \beta_{age} \times (ln(M_{LOW,0,2019}) - ln(M_{loc,0,1950})) \quad (5)$$

For each of these counterfactual age specific mortality profile I can construct country level predictions for life expectancy by age. Table 4 shows the relationship between the predictions from Equation (3) and reality. Columns 1 and 2 are for life expectancy at birth. Columns 3 and 4 show life expectancy at age 10. The latter group has the advantage of removing all the life expectancy movements that come directly from the first year of life which is the starting point from the predictions. Keep in mind that these predictions are generated using only the 1950 age mortality profile, the coefficients estimated earlier in the paper, and the actual changes in age 0-1 mortality over time. These predictions capture well over half of the life expectancy change between 1950 and 2019.

Figure 5 shows the accompanying scatter plot. The outliers are, unsurprisingly, countries that were undergoing unusual circumstances in the initial year, like the Koreas. Figure 6 shows the difference between the predictions from Equation (3) (based on the realized path of age 0-1 mortality) and Equation (5) (based on all countries getting to the lowest current age 0-1 mortality level. This represents potential gains to life expectancy that we should expect to see with convergence in age 0-1 mortality to the current frontier.

Table 4: Predicted versus actual Life Expectancy

	(1) Birth	(2) Birth	(3) Age 10	(4) Age 10
Predicted Change	1.687	1.196	1.767	1.181
	(0.151)	(0.307)	(0.172)	(0.364)
Life Expectency in 1950		-0.208		-0.234
		(0.145)		(0.158)
Constant	6.939	22.44	7.020	22.08
	(1.213)	(10.50)	(1.190)	(9.956)
r2	0.734	0.754	0.674	0.704
N	98	98	98	98

Standard errors in parentheses

Regressions include ageXyear, locationXyear, and locationXyearX10yeargroup

Figure 5: Predicted versus actual Life Expectancy at Age 10 Change

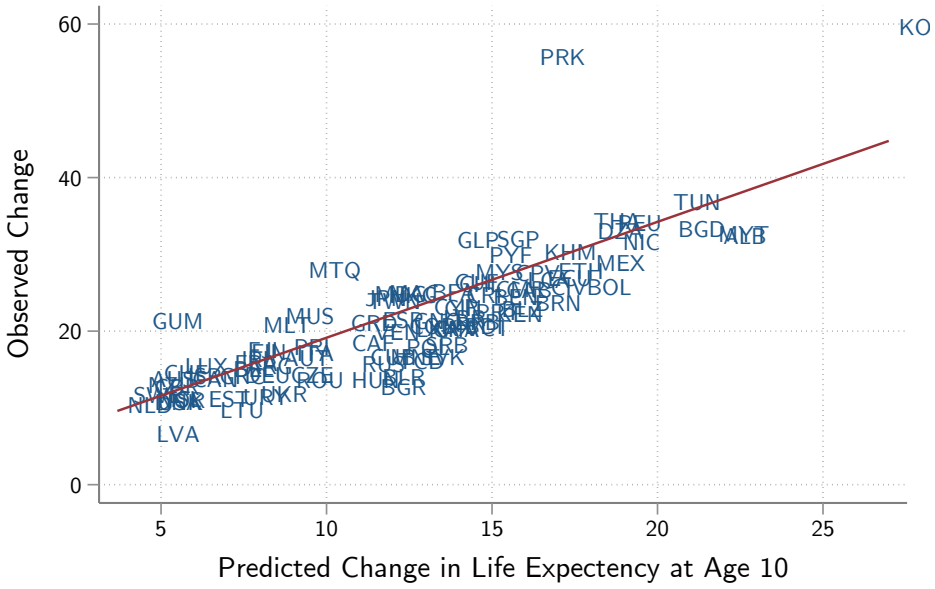
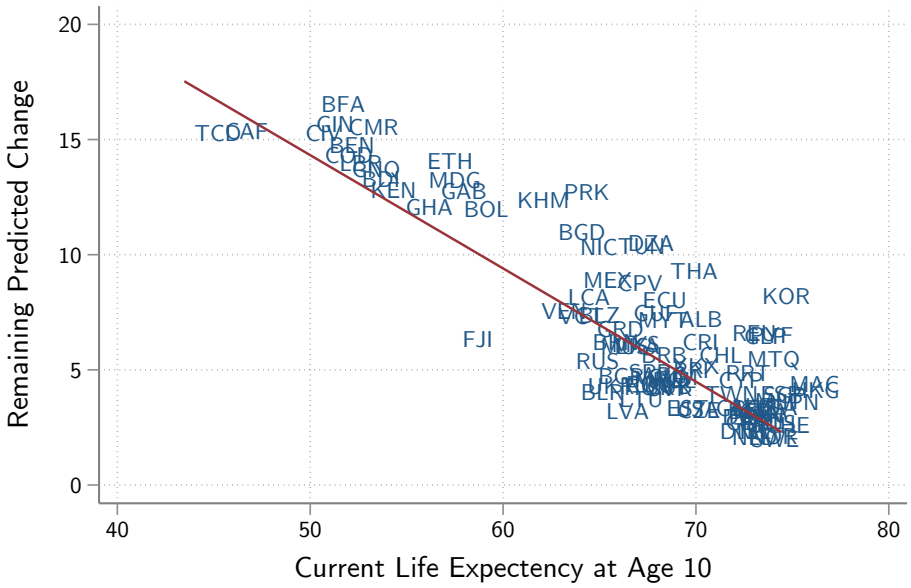


Figure 6: Remaining Gains



References

Almond, Douglas and Janet Currie, "Killing Me Softly: The Fetal Origins Hypothesis," *Journal of Economic Perspectives*, 2011, *25* (3), 153–72.

___ , ___ , and Valentina Duque, "Childhood Circumstances and Adult Outcomes: Act II," *Journal of Economic Literature*, December 2018, *56* (4), 1360–1446.

Barker, D J, "The fetal and infant origins of adult disease.," *BMJ : British Medical Journal*, November 1990, *301* (6761), 1111. 1

Barker, D. J. and C. Osmond, "Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales," *Lancet (London, England)*, May 1986, *1* (8489), 1077–1081. 1

Barker, D. J. P., "Fetal origins of coronary heart disease," *BMJ*, July 1995, *311* (6998), 171–174. Publisher: British Medical Journal Publishing Group Section: Education and debate. 1

Bloom-Feshbach, Kimberly, Wladimir J. Alonso, Vivek Charu, James Tamerius, Lone Simonsen, Mark A. Miller, and Cécile Viboud, "Latitudinal Variations in Seasonal Activity of Influenza and Respiratory Syncytial Virus (RSV): A Global Comparative Review," *PLOS ONE*, February 2013, *8* (2), e54445. Publisher: Public Library of Science. 12

Cannell, J. J., R. Vieth, J. C. Umhau, M. F. Holick, W. B. Grant, S. Madronich, C. F. Garland, and E. Giovannucci, "Epidemic influenza and vitamin D," *Epidemiology & Infection*, December 2006, *134* (6), 1129–1140. 12

Lleras-Muney, Adriana and Flavien Moreau, "A Unified Model of Cohort Mortality," *Demography*, December 2022, *59* (6), 2109–2134.

United Nations, *World Population Prospects 2022: Summary of Results* Statistical Papers - United Nations (Ser. A), Population and Vital Statistics Report, United Nations, August 2022. 2, 4

A Appendix

Table 5 deals with the within country age profile in several different ways. Column 1 omits any controls for the country age profile (but still controls for ageXyear and countryXyear). Column 2 includes country X age controls. Full country age year controls are not possible, but I can allow for the country level age mortality profile to change over time in multiple year groups. Column 3 adds country x age x 5year controls. Column 4 adds country x age x 2year controls. While accounting for differences in the age mortality profile across counties is important, allowing it to vary over time does not seem to significantly change the results.

Table 6 looks at differences in impacts by gender. There are no significant differences by gender.

Table 5: Controlling for locXageXtime

	(1) None	(2) locXage	(3) locXageXyear10	(4) locXageXyear5	(5) locXageXyear2
ln(cohort mx 0-1)	0.699	0.238	0.267	0.257	0.307
	(0.0876)	(0.0370)	(0.0402)	(0.0594)	(0.0817)
R2	0.951	0.992	0.996	0.997	0.998
Within R2	0.231	0.0536	0.0381	0.0215	0.0179
Observations	236572	236474	235886	235200	233240
Countries	98	98	98	98	98
Years	68	68	68	68	68
Cohorts	69	69	69	69	69

Standard errors in parentheses
Regressions include ageXyear, locationXyear, and locationXyearX10yeargroup

Table 6: Gender Differences

	(1) All	(2) Female	(3) Male
ln(cohort mx 0-1)	0.267		
	(0.0402)		
ln(cohort mxf 0-1)		0.320	
		(0.0621)	
ln(cohort mxm 0-1)			0.281
			(0.0454)
R2	0.996	0.991	0.994
Within R2	0.0381	0.0282	0.0285
Observations	235886	235886	235886
Countries	98	98	98
Years	68	68	68
Cohorts	69	69	69

Standard errors in parentheses
Regressions include ageXyear, locationXyear, and locationXyearX10yeargroup