Web Appendix
Index of Web Appendix

Page 2: Footnote 2 (also discussed in Page 17): 1980 Census with other outcomes
Page 3: Footnote 8: correlation matrix between health/education expenditure and pandemic severity
Page 3: Footnote 16: raw MMR figure. Maternal Mortality Rates over time.
Page 4-6: Footnote 21/Footnote 23:
- 1920 birth cohort dummies are removed from the regressions
- change cohort windows to 1915-1927 and 1917-1925
- we add quadratic time trends;
- calendar of months fixed effects are included

Page 7-8: Footnote 22: trimester analysis
Page 9: Footnote 24: probit regression results
Page 9-11 : Page 7 footnote 7: checking birth rates and birth frequencies by the father's residency and field of employment
Page 12:
- Probit Estimates of Effect of Maternal Mortality Rate (MMR) on Educational Attainment 1917-1922
- Using contemporaneous MMR

Page 13-14 :
- Table 3 + region-year level government expenditure on 7-year lagged education, public health and GDP.
- Table 3, but quarter-by-quarter comparison

Page 15-16 :
- Replication of Table 4 but excluding 1920 birth cohort
- Various Specifications of Table 4 Row 1
  - Regression with only weighted MMR, no other control, no bootstrap/cluster
  - Regression with only weighted MMR, gender, IMR, no other control, no bootstrap/cluster
  - Regression with only weighted MMR, gender, IMR, regional dummies, no bootstrap/cluster
  - Regression with only weighted MMR, gender, IMR, regional dummies, region-specific time trend, no bootstrap/cluster

Footnote 2:

Table: 1980 Census, Other Outcomes			
	(1)	(2)	(3)
VARIABLES	Housing Condition	Log (House Size)	Currently Employed
Panel A: All			
Mean (Dep. Variable)	3.14	5.47	0.40
1919 Birth Cohort	-0.0111	0.0016	-0.0217*
	[0.1440]	[0.7246]	[0.0803]
1920 Birth Cohort	-0.0031	0.0043	-0.0199**
	[0.4110]	[0.5269]	[0.0192]
Panel B: Male Only			
Mean (Dep. Variable)	3.15	5.46	0.70
1919 Birth Cohort	-0.0036	0.0060	-0.0331
	[0.6900]	[0.1750]	[0.1208]
1920 Birth Cohort	-0.0002	0.0051	-0.0310**
	[0.9732]	[0.6112]	[0.0380]
Panel C: Female Only			
Mean (Dep. Variable)	3.14	5.47	0.13
1919 Birth Cohort	-0.0179*	-0.0025	-0.0111***
	[0.0649]	[0.6948]	[0.0095]
1920 Birth Cohort	-0.0057	0.0035	-0.0088**
	[0.1948]	[0.5125]	[0.0204]

Notes: P-values from clustering at regional level are in brackets. * Significant at 10% level, ** at 5% level, and *** at 1% level. Regression specification is the same as Equation 1. Birth cohorts included are 1916–1926. Infant mortality rate, region dummies, and region-specific time trends are included. Panel A also include a gender dummy. The dependent variable in Column 1 is an index measuring one's living condition (i.e., whether residence has separate versus shared bathroom, kitchen, living room and piped water), ranging from 1 to 4. The higher index indicated a better living condition.

Footnote 8:

```
             |          ir       mmr   education Sanitation GDP
-------------+------------------------------------------------
          ir |     1.0000
         mmr |     0.5109    1.0000
   Education |     0.6471    0.5792    1.0000
  Sanitation |     0.2093    0.2375    0.7428    1.0000
         GDP |     0.7229    0.5849    0.9205    0.5920    1.0000
```

Ir = influenza death rate
Mmr = maternal mortality rates
Education = Total Education Expenditure for a given region-year (7-years lag)
Sanitation = Total Sanitation Expenditure for a given region-year (contemporaneous)
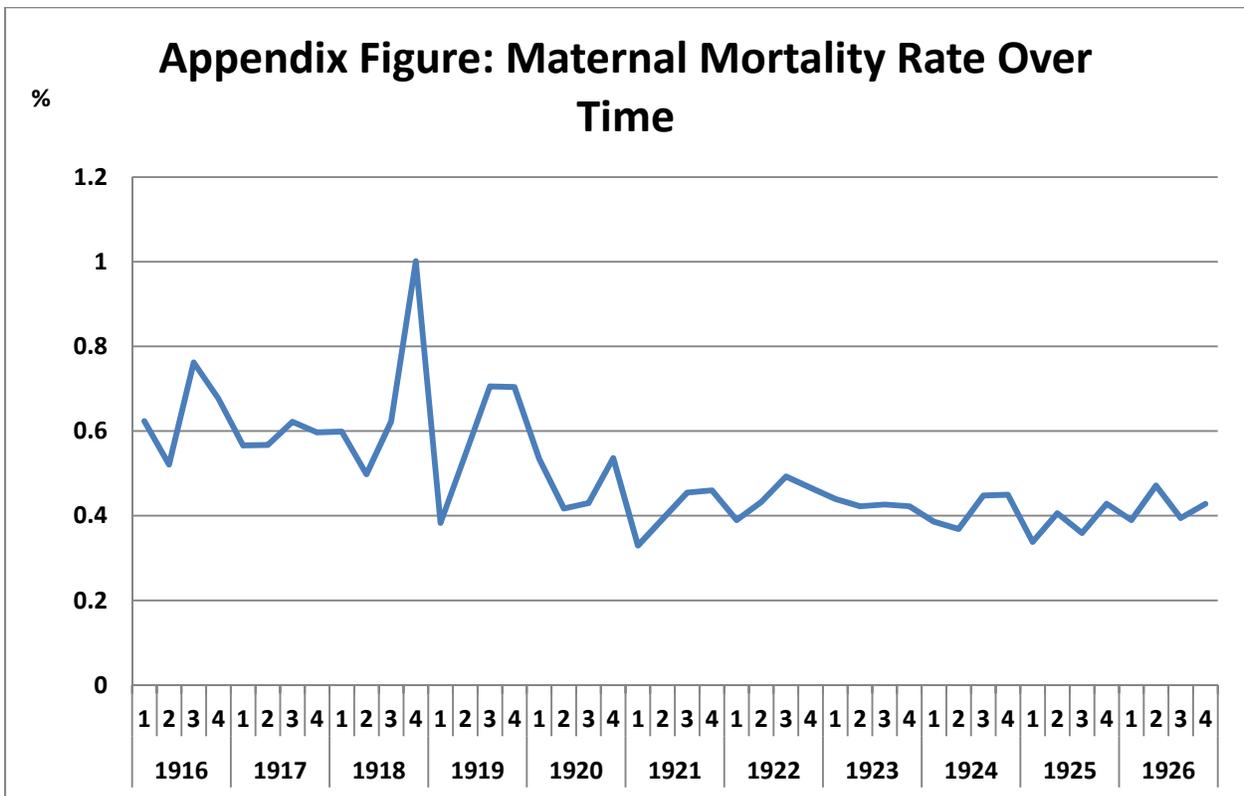GDP = Total GDP for a given region-year (contemporaneous)

Footnote 16:

Figure above shows an overall declining trend of maternal mortality rates with a few spikes during 1916–1926. It shows that during the final quarter of 1918, the probability of women dying from childbirth rose to 1 percent (November: 1.4 percent; December: 0.95 percent) in 1918, compared to fluctuations between

0.4 percent and 0.8 percent during other periods.[1]

Footnote 21: Dropping 1920 cohort dummy.

Appendix Table: Departure in Education Outcomes of 1919 Birth Cohort, 1916-1926

VARIABLES	(1) Years of Schooling	(2) Elementary School	(3) Middle School	(4) High School
1919 Birth Cohort	-0.073***	-0.007***	-0.005***	-0.005***
	[0.000]	[0.000]	[0.000]	[0.000]
Observations	870,468	870,468	870,468	870,468
R-squared	0.170	0.175	0.047	0.027

Notes: Wild bootstrap p-values with 500 repetitions and p-values from clustering at regional level are in brackets. * significant at 10% level, ** at 5% level and *** at 1% level based on wild bootstrap p-values. Birth cohorts included are 1916-1626. All models allow for correlation among observations within the same region. Gender, region dummies, region specific time trends, infant mortality rates are included.

---

[1] The Japanese colonial government invested much effort in collecting detailed records of causes of death, including those due to various infectious diseases. Therefore, we can identify the timing of each outbreak of disease and the severely infected areas. The spike in 1916 was a malaria outbreak that caused 11,000 deaths (0.3 percent of the total population) and the spike in summer of 1919 was a cholera outbreak during which 3,826 contracted cholera and 2,693 died (0.07 percent of the total population) (Sun 2010). The 1916 birth cohort would be excluded from the analysis. On the other hand, another cholera outbreak took place between two flu pandemics. The cholera outbreak has a few characteristics: first, the mortality rate for the 1919 cholera outbreak was high at 71 percent, meaning that there were fewer than 1,200 survivors in the total population of 3 million. While cholera outbreak increases maternal mortality rates, it could be difficult to detect the long-term effect in the overall population, which would possibly bias our estimates downward. Second, more than 85 percent of 1919 cholera cases were in the Taipei, Tainan, and Taidong regions, so in the robustness check section, we could exclude these regions, and still find similar results.

- **Including Month Fixed Effects**

Appendix Table: Departure in Education Outcomes of 1919 Birth Cohort, 1916-1926
Including Month Fixed Effects

VARIABLES	(1) Years of Schooling	(2) Elementary School	(3) Middle School	(4) High School
Mean (Dependent Variable)	*3.32*	*46.98%*	*9.92%*	*5.23%*
1919 Birth Cohort	-0.084***	-0.008***	-0.006***	-0.005***
	[0.000]	[0.002]	[0.000]	[0.000]
1920 Birth Cohort	-0.056***	-0.006***	-0.004***	-0.003***
	[0.002]	[0.007]	[0.002]	[0.000]
Observations	870,468	870,468	870,468	870,468
R-squared	0.170	0.175	0.047	0.027

Notes: p-values from clustering at regional level are in brackets. * significant at 10% level, ** at 5% level and *** at 1% level based on wild bootstrap p-values.   Birth cohorts included are 1916-1626. All models allow for correlation among observations within the same region. Gender, region dummies, region specific time trends are included.

- **Including Quadratic Trends**

Table: Departure in Education Outcomes of 1919 and 1920 Birth Cohorts with
Quadratic Time Trends

VARIABLES	(1) Years of Schooling	(2) Elementary School	(3) Middle School	(4) High School
1919 Birth Cohort	-0.063***	-0.005***	-0.004***	-0.005***
	[0.001]	[0.006]	[0.001]	[0.000]
1920 Birth Cohort	-0.033**	-0.003*	-0.002***	-0.002***
	[0.027]	[0.094]	[0.002]	[0.005]
Observations	870,468	870,468	870,468	870,468
R-squared	0.169	0.175	0.047	0.027

Notes: Standard errors are clustered at the regional level. P-values from clustering at regional level are in brackets. * significant at 10% level, ** at 5% level and *** at 1% level. Birth cohorts included are 1916–1926. All models allow for correlation among observations within the same region. Gender, region dummies, year-of-birth (yob), $(yob)^2$ are included.

Footnote 21/Footnote 23: Changing Birth Cohorts 1915-1927; 1927-1925

Table: Departure in Education Outcomes of 1919 and 1920 Birth Cohorts, 1915-1927

VARIABLES	(1) Years of Schooling	(2) Elementary School	(3) Middle School	(4) High School
1919 Birth Cohort	-0.010***	-0.003**	-0.007***	-0.006***
	(0.0204)	(0.001)	(0.001)	(0.001)
1920 Birth Cohort	-0.075***	-0.004**	-0.005***	-0.003***
	(0.0152)	(0.002)	(0.001)	(0.001)
Observations	1,045,946	1,045,946	1,045,946	1,045,946
R-squared	0.177	0.097	0.049	0.028

Notes: Standard errors are clustered at the regional level. P-values from clustering at regional level are in brackets. * significant at 10% level, ** at 5% level and *** at 1% level. Birth cohorts included are 1915–1927. All models allow for correlation among observations within the same region. Gender, region dummies, year-of-birth, regional time trends are included.

Table: Departure in Education Outcomes of 1919 and 1920 Birth Cohorts, 1917-1925

VARIABLES	(1) Years of Schooling	(2) Elementary School	(3) Middle School	(4) High School
1919 Birth Cohort	-0.049**	0.001	-0.004	-0.004***
	(0.0171)	(0.001)	(0.001)	(0.001)
1920 Birth Cohort	-0.035**	0.000	-0.003***	-0.002***
	(0.0137)	(0.002)	(0.001)	(0.000)
Observations	714,329	714,329	714,329	714,329
R-squared	0.164	0.091	0.046	0.026

Notes: Standard errors are clustered at the regional level. P-values from clustering at regional level are in brackets. * significant at 10% level, ** at 5% level and *** at 1% level. Birth cohorts included are 1917–1925. All models allow for correlation among observations within the same region. Gender, region dummies, year-of-birth, regional time trends are included.

Footnote 22: For trimester regression

The specification is as follows:

$$y_{ijt} = \alpha + \beta_1 * I(first\ trimester\ in\ pandemic) * I(wave1)$$
$$+\beta_2 * I(second\ trimester\ in\ pandemic) * I(wave1)$$
$$+\beta_3 * I(third\ trimester\ in\ pandemic) * I(wave1)$$
$$+\beta_4 * I(first\ trimester\ in\ pandemic) * I(wave2)$$
$$+\beta_5 * I(second\ trimester\ in\ pandemic) * I(wave2)$$
$$+\beta_6 * I(second\ trimester\ in\ pandemic) * I(wave2)$$
$$+ X_i + \sum_j Region\ Fixed\ Effects_j + \sum_j Region_j * Time_t + \varepsilon_{ijt}$$

----- Equation (2)

I(.) is an indicator function whether individual i in region j born in year t had experienced Pandemic in the first, second, or third trimester, and whether the exposure was during the first wave or second wave of Pandemic. X_i includes gender, region-specific infant mortality rates, and birth-quarter fixed effects. We code the wave 1 dummy variable equals to one if one was in utero between October-December, 1918; wave 2 dummy variable equals to one if one was in utero between January–February 1920. Each of the trimester dummy equals to one if two out of the three months in the trimester were during the peak of pandemic (November–December 1918 or January–February 1920); the dummy would equal zero if none of the three months in the trimester was during the peak. The results are presented in Table 3. We find that in utero exposure to the first wave have larger negative effects than the exposure to the second wave across all education outcomes.

Table: The Impact of In Utero Exposure to Influenza by Trimester by Wave	
Dep Variable: Years of Education	
1st trimester influenza exposure during wave 1	-0.080
	[0.053]
2nd trimester influenza exposure during wave 1	-0.055**
	[0.022]
3rd trimester influenza exposure during wave 1	-0.091*
	[0.042]
1st trimester influenza exposure during wave 2	-0.017
	[0.036]
2nd trimester influenza exposure during wave 2	-0.051
	[0.030]
3rd trimester influenza exposure during wave 2	-0.030*
	[0.016]
Observations	399,818
R-squared	0.150

Standard errors clustered at the region level are presented in brackets. * significant at 10% level, ** at 5% level and *** at 1%. We code each of the trimester as 1 if one experiences 2 out of 3 months of trimester during the peak of pandemic (defined as Nov/Dec 1918 or Jan/Feb 1920), 0 if zero month out of 3 months of trimester during the peak of pandemic Those who experience only 1 month of trimester during the peak of pandemic are excluded from the regression. Wave 1 is defined as Nov/Dec 1918. Wave 2 is defined as Jan/Feb 1920. Infant mortality rate, gender, region dummies and region specific time trends are also included in the regression.

Appendix: Effect of Maternal Mortality Rate on Disease Prevalence in the 1989 Elderly Survey

	Kidney Disease	Vertigo (Dizzy)	Circulatory disease	Respiratory disease	Glaucoma	Diabetes
	(1)	(2)	(3)	(4)	(5)	(6)
Mean (Dependent Variable)	*0.047*	*0.271*	*0.163*	*0.146*	*0.014*	*0.051*
Panel B: Probit Model						
weighted maternal mortality rates	0.048***	0.283	0.170*	0.176**	0.007***	0.092**
p-value from clustering	(0.002)	(0.149)	(0.079)	(0.038)	(0.000)	(0.014)

Notes: N=1836. p-values from clustering standard errors at regional level are in parentheses. Marginal effects from probit model are reported. * significant at 10% level, ** at 5% level, and *** at 1% level based on p-values. Birthyear ranges from 1916 to 1926. Infant mortality rate, gender, region dummies, and region-specific time trends are also included in all regressions.

-

Probit Model of Father's Literacy and Occupation using
1917-1926 Birth Cohorts

Dependent Variable	(1) Father is Farmer	(2) Father's Illiterate
1919 Birth Cohort	-0.03	-0.04
	(0.16)	(0.09)
1920 Birth Cohort	-0.12	-0.07
	(0.09)	(0.09)
Constant	9.38	44.89*
	(28.20)	(25.16)
Observations	1,767	1,653

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

- Deviation in birth rates of Japanese residents during 1919/1920

Appendix Table: Deviation in Birth Rates By Origin During Flu Pandemic

Dependent Variable: Birth Rates	
Japan=1	0.628***
	(0.117)
Japan * Linear Time Trend	-0.000332***
	(6.10e-05)
Linear Time Trend	0.000203***
	(4.82e-05)
Year=1919	-0.00278
	(0.00256)
Year=1920	-0.00191
	(0.00256)
Japan * (Year=1919)	0.00456
	(0.00362)
Japan * (Year=1920)	0.00331
	(0.00361)
Constant	-0.346***
	(0.0926)
Observations	70
R-squared	0.853

Data on number of births is available from 1906 to 1943. Combining this data with total population allows us to construct birth rates. Standard errors in parentheses. The default group is birth rates of Taiwanese.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

[This data was collected by Japanese Colonial Government, and it includes both live and still births.]

Appendix Table: Deviation in Birth Frequencies by Father's Field of Employment during Flu Pandemic

	(1)	(2)	(3)
Dep Variable: Birth Frequencies			
1919	-2,005	-1,981	-1,981
	(11,941)	(11,961)	(2,142)
1920	-1,876	-1,857	-1,857
	(11,928)	(11,947)	(2,139)
Government	-12,815**	-12,014	3,485
	(6,188)	(12,573)	(2,450)
Government * 1919	1,019	784.6	784.6
	(37,638)	(37,823)	(6,772)
Government * 1920	895.4	701.8	701.8
	(37,638)	(37,780)	(6,765)
Linear Time Trend	X	X	X
Government * Linear Time Trend		X	X
Fields of Employment Fixed Effect			X
Constant	12,186***	12,105***	-3,393***
	(3,817)	(3,976)	(1,200)
Observations	380	380	380
R-squared	0.029	0.029	0.970

Standard errors in parentheses. Data is available between 1906 and 1943. Fields of employment include agriculture, transportation, fishery, mining, industry, commercial, government, others, unemployed and housework. 1919/1920 indicates dummies for birth frequencies in 1919/1920 respectively. Government is a dummy indicating whether father works for government. Shaded areas are the coefficients of interest, indicating the deviation in birth frequencies for civil servants.

*** p<0.01, ** p<0.05, * p<0.1

Appendix Table : Probit Estimates of Effect of Maternal
Mortality Rate on Educational Attainment 1917-1922

	Dependent variable			
	Elementary School	Middle School	High School	College
weighted maternal mortality rates	-0.29 (0.18)	-0.28* (0.15)	-0.39** (0.18)	-0.43** (0.18)

Note: Same specification and covariates as Table 4, Specification 2. Standard errors clustered at the regional level and reported in parentheses. ***, **, and * are significant at 1%, 5%, 10% respectively.

o **Contemporaneous MMR:**

Appendix Table : Estimates of Effect of Maternal
Mortality Rate on Educational Attainment 1917-1922

	Dependent variable			
	Yrs of education	Elementary School	Middle School	High School
contemporaneous MMR	-0.52 (0.43)	-0.06 (0.05)	-0.02 (0.02)	-0.02 (0.02)

Note: Same specification and covariates as Table 4, Specification 2 except the MMR is now defined as the contemporaneous MMR from given region-year. Standard errors clustered at the regional level and reported in parentheses. ***, **, and * are significant at 1%, 5%, 10% respectively.

(Lagged MMR is reported in Table 4 Specification 3)

- Table 3 + region-year level government expenditure on 7-year lagged education, public health and GDP.

<div align="center">Appendix Table 3: Same Specification As Table 3
with Government Health and Education Exp</div>

VARIABLES	Years of Schooling	Elementary School	Middle School	High School
1919 Birth Cohort	-0.110***	-0.0110**	-0.00667***	-0.00661***
p-value from wild bootstrap	[0.000]	[0.024]	[0.000]	[0.000]
1920 Birth Cohort	0.0147	0.00162	-0.00102	0.000626
p-value from wild bootstrap	[0.812]	[0.768]	[0.840]	[0.964]
Observation	870468	870468	870468	870468

Note: N=870,468. Each column represents results from a regression. Wild bootstrap p-values with 500 repetitions and p-values from clustering at regional level are reported in brackets. All models allows for correlation among observations within the same region. Infant mortality rate, gender, education expenditure (7-years lag), sanitation expenditure, region dummies and region-specific time trends are also included in all regressions.

- Table 3 but in birth Quarter Comparison

Appendix Table: Departure in Education Outcomes from 1917-1926 Birth Cohorts

	Dependent variable			
	Yrs of education	Elementary School	Middle School	High School
q_19184	-0.0039	0.0014	-0.0021	-0.0020
	(0.0373)	(0.0059)	(0.0022)	(0.0018)
q_19191	-0.0630**	-0.0041	-0.0064**	-0.0052***
	(0.0261)	(0.0028)	(0.0022)	(0.0016)
q_19192	-0.1091***	-0.0093***	-0.0077***	-0.0066***
	(0.0159)	(0.0025)	(0.0014)	(0.0010)
q_19193	-0.0370	-0.0037	-0.0025	-0.0031*
	(0.0367)	(0.0043)	(0.0025)	(0.0015)
q_19194	-0.0567	-0.0038	-0.0051**	-0.0063***
	(0.0336)	(0.0039)	(0.0023)	(0.0020)
q_19201	-0.0620**	-0.0067*	-0.0043**	-0.0028**
	(0.0230)	(0.0032)	(0.0015)	(0.0010)
q_19202	-0.0756	-0.0081	-0.0072***	-0.0021
	(0.0547)	(0.0082)	(0.0022)	(0.0013)
q_19203	-0.0274	-0.0017	-0.0045	-0.0025
	(0.0447)	(0.0054)	(0.0030)	(0.0015)
Constant	-329.1337***	-45.0720***	-12.1505***	-5.8007***
	(4.6927)	(0.7693)	(0.1145)	(0.1090)
Observations	870,468	870,468	870,468	870,468
R-squared	0.1698	0.1752	0.0475	0.0274

Note: Each of the coefficient represent deviation in years of education for a given birth year-quarter cohort relative to those who were born between 1917 to 1926. All models allows for correlation among observations within the same region. Infant mortality rate, female dummy, birth quarter dummies, region dummies and region-specific time trends are included. The highlighted areas correspond to 9-months delay of the two spikes occurred in Figure 2b in 1918 4th quarters and 1919 4th quarter.

Appendix Table: Replication of Table4 excluding 1920 birth cohort

	Years of Schooling	Elementary School	Middle School	High School
(1): Estimation of Equation 3				
weighted maternal mortality rates	-1.823***	-0.237***	-0.075***	-0.046***
p-value from cluster	[0.000]	[0.000]	[0.000]	[0.000]
p-value from wild bootstrap	[0.008]	[0.008]	[0.008]	[0.008]
(2): (1) + Region-specific time trends				
weighted maternal mortality rates	-0.886	-0.098	-0.047**	-0.038*
p-value from cluster	[0.157]	[0.194]	[0.107]	[0.085]
p-value from wild bootstrap	[0.108]	[0.204]	[0.048]	[0.084]
(3): (2) + MMR(t-1)				
MMR (t-1)	-0.921*	-0.116	-0.040*	-0.028*
p-value from cluster	[0.149]	[0.166]	[0.107]	[0.088]
p-value from wild bootstrap	[0.076]	[0.128]	[0.068]	[0.068]
(4): (2) + Government Sanitation Expenditure, Education Expenditure and regional GDP per capita				
weighted maternal mortality rates	-0.851	-0.099	-0.042*	-0.032*
p-value from cluster	[0.196]	[0.219]	[0.167]	[0.141]
p-value from wild bootstrap	[0.132]	[0.228]	[0.052]	[0.068]
(5): (2) + 9-Month Average MMR				
9-month moving average MMR	-1.581	-0.180	-0.080**	-0.064***
p-value from cluster	[0.152]	[0.199]	[0.095]	[0.039]
p-value from wild bootstrap	[0.328]	[0.624]	[0.028]	[0.000]

Note: N=371,530. Each coefficient represent a result from a regression. Wild bootstrap p-values with 500 repetitions and p-values from clustering at regional level are in brackets. All models allow for correlation among observations within the same region. Infant mortality rate, gender and region dummies are also included in all regressions.

Appendix Table : Estimates of Effect of Maternal Mortality Rate on Educational Attainment 1917-1922

	Dependent variable			
	Yrs of education	Elementary School	Middle School	High School
(1): Only control for MMR				
Weighted MMR	-1.58***	-0.23***	-0.03***	-0.03***
	(0.04)	(0.00)	(0.00)	(0.00)
(2): (1) + Female Dummy + Infant Mortality Rate				
Weighted MMR	-1.43***	-0.21***	-0.02***	-0.02***
	(0.04)	(0.00)	(0.00)	(0.00)
(3): (2) + Regional Dummy				
Weighted MMR	-1.77***	-0.23***	-0.07***	-0.04***
	(0.05)	(0.01)	(0.00)	(0.00)
(4): (3) + Region Dummy and Region Specific Time Trend				
Weighted MMR	-0.87***	-0.10***	-0.04***	-0.04***
	(0.08)	(0.01)	(0.01)	(0.00)

Each coefficient is from a separate regression. (3) is the same specification as Table 4 Spec 1, but standard errors are not clustered/bootstrapped. (4) is the same specification as Table 4 Spec (2), but the standard errors are not clustered/bootstrapped