Economic impacts of reducing the infectious disease burden in the US: Evidence using population-level differences in genetic resistance

Justin Cook and Jason Fletcher
UC-Merced and UW-Madison

IGSS 2015

Motivating question

Do population wide health improvements lead to improvements in aggregate economic well being?

Background (I)

Question and evidence initially posed by Acemoglu and Johnson (2007, JPE).

- Attempt to disentangle the causative relationship by using:
  1. The discovery of a number of medical innovations (c. 1950).
  2. The pre-innovation mortality rate from infectious disease (intensity of this treatment).
     - Countries with higher initial mortality rates will benefit more from the innovations.
     - Likely tied to other factors of economic growth–i.e., does not satisfy the exclusion restriction.
- Instrument change in life expectancy between 1940 and 1980 to predict its causative effect on total output, population, and output per capita.

Background (II)

- Find positive effects on population, an insignificant effect on output, and a negative effect on output per capita.
- Malthusian response in which technology improvement outstrips any positive economic effects.

	B. Dependent Variable: Log per Capita GDP						
Log life expectancy	−1.32	−1.51	−2.35	−2.70	−1.64	−1.59	−1.21
	(.56)	(.57)	(1.13)	(1.40)	(.77)	(1.22)	(.52)
Postyear dummy × institutions or initial log per capita GDP					−.049	−.073	
					(.060)	(.278)	
Number of countries	47	47	36	36	47	47	47

- This result has recently been replicated within the US (Hansen 2014, JDE).

What we do (I)

Re-estimate the proposed relationship of AJ using an instrument that is plausibly more exogenous: a measure of genetic resistance to infectious disease.

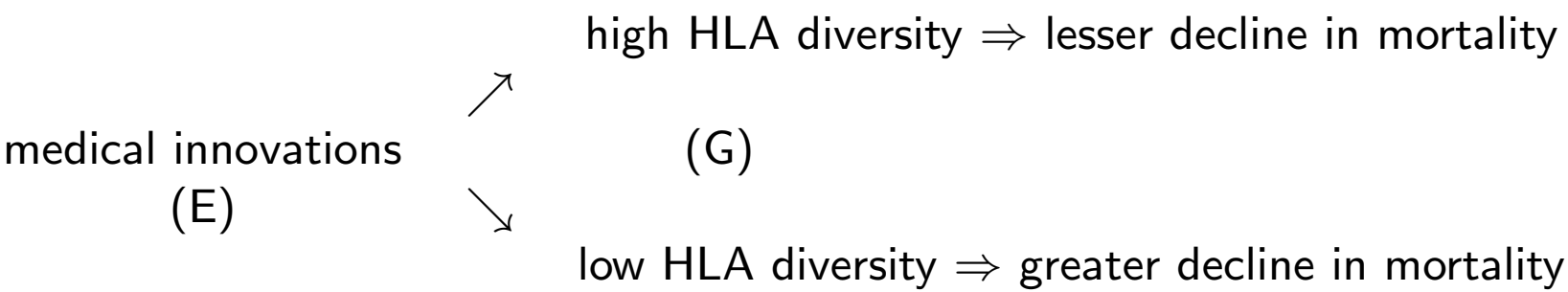
- Our measure of genetic resistance comes from Cook (2015, ReStat).
  - Genetic diversity within the human leukocyte antigen (HLA) system favorable for resistance.
- To avoid confounders, improve data quality, etc., we focus our analysis on states within the US.

What we do (II)

To do so, we will:

1. Convert the country-level measure from Cook to the state level.
2. Use this state-level measure as a measure of intensity of treatment.

A gene-environment interaction:



What we find

↓ing mortality from infectious disease or ↑ing life expectancy led to:

- $\Rightarrow$ ↑ed growth in population
- $\Rightarrow$ ↑ed growth in output
- $\Rightarrow$ ↑ed growth in output per capita
- $\Rightarrow$ ↑ed growth in general equilibria effects
  - i.e., years of schooling& labor force participation.

These findings counter those of AJ and Hansen.

HLA Heterozygosity

Our primary measure of *population* resistance to infectious disease is HLA heterozygosity.

- Measure of genetic diversity composed solely of genes (SNPs) comprising the HLA system.
  - 156 SNPs for 51 ethnicities; ALFRED.
  - Ethnic data are aggregated to country level by Cook.
- Diversity within the HLA system is associated with diversity of immune response (Doherty and Zinkernagel 1975).
  - Diverse immune responses slow strains that are able to overcome a common response.
  - More than 2x likely to die if obtain measles from a relative (Garenne and Aaby 1992).

Conversion to US States

Self-reported ancestry is first found for the 1980 Census.

- We match this ancestry measure to country/ethnicity of Cook. Also for Ashraf and Galor's overall diversity.
- We don't want 1980 populations, rather we want the population in the ex-ante period, 1940.
  - Use state of birth for those 40 or older in 1980.
- Create representative state-level HLA heterozygosity score from the weighted average of the ancestral composition, representative of 1940s population.

Other Variables (All by Decade, 1940-2000)

Mortality from infectious disease.

- Sum of mortality from influenza, pneumonia, tuberculosis, syphilis, typhoid, dysentery, diphtheria, whooping cough, meningococcal infections, polio, measles, and other deaths from infectious disease.
- From National Vital Statistics.

Life Expectancy at birth

- From National Vital Statistics.

State-Level Economic Outcomes.

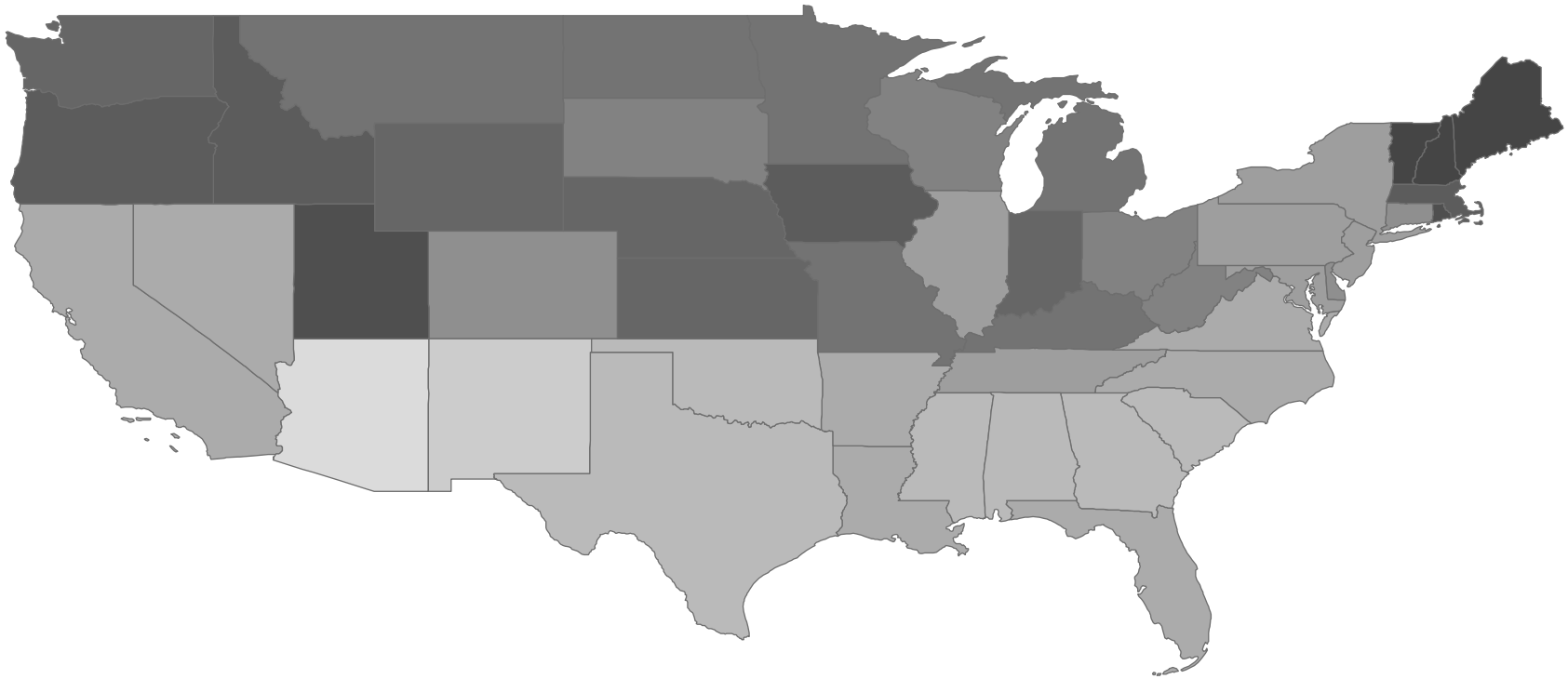
- Includes population, labor force, years of schooling, and income.
- From Turner et al. (2007, JEG).
- Ranges from 1840 to 2000.

Table 1. Summary Statistics

Variable	Obs.	Mean	Standard Deviation
State HLA Heterozygosity, 1940	48	0.3379	0.0053
By Region:			
South	16	0.3353	0.003
Northeast	9	0.3416	0.0032
Midwest	12	0.3401	0.0009
West	11	0.3365	0.0087
Mortality Rate from Infectious Disease, 1940	48	93.21	39.95
By Region, 1940:			
South	16	124.55	17.49
Northeast	9	63.76	9.28
Midwest	12	66.04	17.26
West	11	101.37	57.73
By Year:			
1940	48	93.21	39.95
1950	48	65.28	18.67
1960	48	48.65	8.60
1970	48	39.31	7.37
1980	48	31.30	5.39
1990	48	49.74	9.29
2000	48	42.54	10.13
Life Expectancy at Birth, 1940	48	64.07	2.40
By Region, 1940:			
South	16	62.01	1.56
Northeast	9	64.89	0.65
Midwest	12	66.36	1.40
West	11	63.90	2.68
By Year:			
1940	48	64.07	2.40
1950	48	68.51	1.47
1960	48	69.97	1.26
1970	48	70.87	1.29
1980	48	74.06	1.12
1990	48	75.62	1.26
2000	48	77.21	1.43

Summary Statistics: State HLA Heterozygosity

State HLA Heterozygosity, 1940	48	0.3379	0.0053
By Region:			
South	16	0.3353	0.003
Northeast	9	0.3416	0.0032
Midwest	12	0.3401	0.0009
West	11	0.3365	0.0087



Summary Statistics: Mortality from Infectious Disease

Mortality Rate from Infectious Disease, 1940	48	93.21	39.95
By Region, 1940:			
South	16	124.55	17.49
Northeast	9	63.76	9.28
Midwest	12	66.04	17.26
West	11	101.37	57.73
By Year:			
1940	48	93.21	39.95
1950	48	65.28	18.67
1960	48	48.65	8.60
1970	48	39.31	7.37
1980	48	31.30	5.39
1990	48	49.74	9.29
2000	48	42.54	10.13

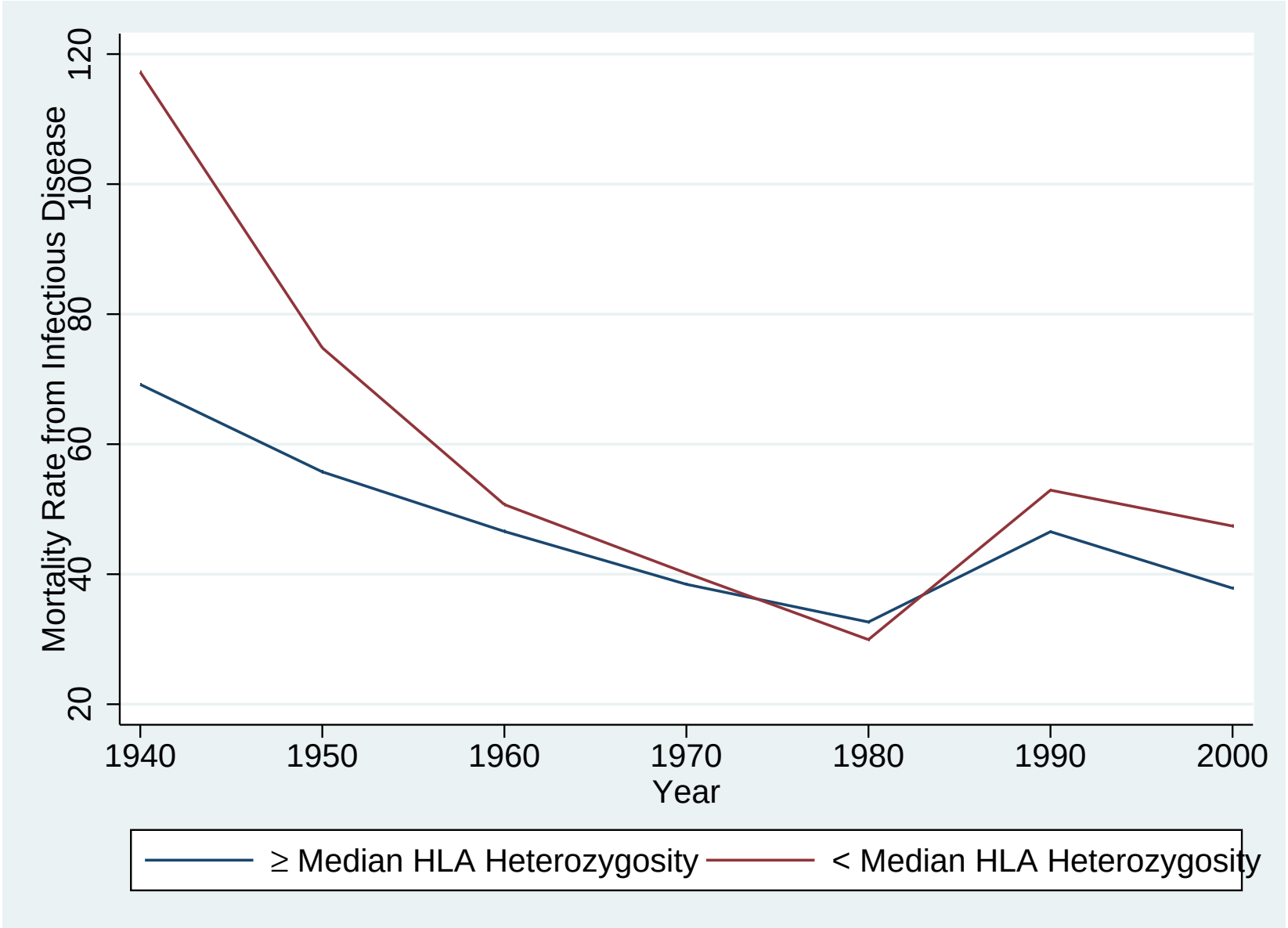
Summary Statistics: Life Expectancy at Birth

Life Expectancy at Birth, 1940	48	64.07	2.40
By Region, 1940:			
South	16	62.01	1.56
Northeast	9	64.89	0.65
Midwest	12	66.36	1.40
West	11	63.90	2.68
By Year:			
1940	48	64.07	2.40
1950	48	68.51	1.47
1960	48	69.97	1.26
1970	48	70.87	1.29
1980	48	74.06	1.12
1990	48	75.62	1.26
2000	48	77.21	1.43

Flexible First Stage Relationship

Table 2. Year by year effects of HLA heterozygosity on Infectious Disease Mortality

Dependent Variable: ln Mortality Rate from Infectious Disease								
Year:	1940	1950	1960	1970	1980	1990	2000	1990-2000
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln HLA Heterozygosity	-11.65***	-5.37**	0.54	0.30	1.69	-4.03	-7.02**	-5.53**
	(2.28)	(2.30)	(2.15)	(2.38)	(1.49)	(2.78)	(2.98)	(2.19)
Controls								
Region	Y	Y	Y	Y	Y	Y	Y	Y
Observations	48	48	48	48	48	48	48	96
R Sqr.	0.80	0.66	0.34	0.37	0.41	0.33	0.53	0.35


Mortality Rate from Infectious Disease
120
100
80
60
40
20
1940
1950
1960
1970
1980
1990
2000
Year
≥ Median HLA Heterozygosity
< Median HLA Heterozygosity

Placebo Test of First Stage

Table 3. Placebo Test: Year by year effects of HLA heterozygosity on All Cause Mortality

Dependent Variable: ln Mortality Rate from All Causes (excluding Infectious Disease)								
Year:	1940	1950	1960	1970	1980	1990	2000	1990-2000
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ln HLA Heterozygosity	2.50	1.44	-0.38	1.67	-0.39	-0.28	0.30	0.01
	(5.43)	(1.16)	(3.07)	(1.62)	(1.39)	(1.62)	(1.75)	(1.17)
Controls								
Region	Y	Y	Y	Y	Y	Y	Y	Y
Observations	48	48	48	48	48	48	48	96
R Sqr.	0.15	0.49	0.20	0.34	0.36	0.40	0.39	0.39

Baseline First Stage Diff-in-Diff

Estimating Equation:

$$x_{it} = \alpha \ln HLA\ het_i \times I_t^{Post} + \theta_{it} \boldsymbol{X}'_{it} + \sum_{j=1940}^{2000} \boldsymbol{X}'_{\boldsymbol{i}} \boldsymbol{I}_t^j \varphi_j + \sum_c \gamma_c I_i^c + \sum_{j=1940}^{2000} \rho_j I_t^j + \varepsilon_{it}$$

- *HLA het* < 1, so ln*HLA het* < 0.
- α hypothesized to be > 0.
  - $\uparrow$ HLA $\Rightarrow$ ln HLA $\rightarrow \quad 0 \Rightarrow$ smaller decline in mort.
  - $\downarrow$ HLA $\Rightarrow$ ln HLA $\rightarrow -\infty \Rightarrow$ larger decline in mort.

Baseline First Stage Diff-in-Diff

Table 4. Baseline First Stage

Dependent Variable: ln Mortality Rate from Infectious Disease				
Sample Period:	1940-2000 (1)	1940-2000 (2)	1940-1980 (3)	1940-2000 (4)
ln HLA Het. × Post-1940 Indicator	16.04***	9.25***	11.23***	11.00***
	(1.47)	(2.47)	(2.46)	(2.41)
ln HLA Het. × Post-1980 Indicator				-5.07**
				(2.16)
Controls				
Time Invariant (× Year)				
Overall Genetic Diversity	N	Y	Y	Y
Income per Capita, 1940	N	Y	Y	Y
Urbanization Rate, 1940	N	Y	Y	Y
Time Variant (× Year)				
Percentage of Non-White Population	N	Y	Y	Y
ln All Cause Mortality Rate (Excluding Infectious Disease)	N	Y	Y	Y
Fixed Effects				
State	Y	Y	Y	Y
Year	Y	Y	Y	Y
Region (× Year)	N	Y	Y	Y
Observations	336	336	240	336
R Sqr.	0.80	0.90	0.94	0.91

Extension of Hansen and AJ

	OLS (1)	2SLS (2)	2SLS (3)
Panel A: ln Population, 1940-1980			
ln Life Expectancy	5.99***	9.07***	7.78**
	(1.61)	(2.84)	(2.92)
Baseline Controls	Y	Y	Y
Instruments			
ln Mort. from Infectious Disease, 1940 × Post-1940	N	Y	N
ln HLA Het. × Post-1940	N	N	Y
Observations	240	240	240
R Sqr.	0.88	–	–
First Stage F Stat	–	63.52	44.67

Extension of Hansen and AJ

Panel B: ln Output, 1940-1980			
ln Life Expectancy	6.12***	9.80***	9.98***
	(1.64)	(3.11)	(3.41)
Baseline Controls	Y	Y	Y
Instruments			
ln Mort. from Infectious Disease, 1940 × Post-1940	N	Y	N
ln HLA Het. × Post-1940	N	N	Y
Observations	240	240	240
R Sqr.	0.97	–	–
First Stage F Stat	–	63.52	44.67

Extension of Hansen and AJ

Panel C: ln Output per Capita, 1940-1980			
ln Life Expectancy	0.13	0.73	2.20*
	(0.40)	(1.09)	(1.31)
Baseline Controls	Y	Y	Y
Instruments			
ln Mort. from Infectious Disease, 1940 × Post-1940	N	Y	N
ln HLA Het. × Post-1940	N	N	Y
Observations	240	240	240
R Sqr.	0.99	–	–
First Stage F Stat		63.52	44.67

Conclusion

- HLA diversity has a strong first stage relationship with state-level mortality from infectious disease and life expectancy.
- Following the innovations of 1950, more diverse HLA states had a lower decline in mortality/slower growth in life expectancy; vice versa.
- Using this relationship we causally estimate the effect of improving health conditions on economic outcomes.
    - Contrary to AJ and Hansen, we find beneficial effects of the specified medical innovations.

Thank you.

Panel A: ln Population				
ln Mortality Rate from Infectious Disease	-0.58***	-1.02***	-0.56***	-0.27**
	(0.20)	(0.30)	(0.17)	(0.10)
Baseline Controls	N	Y	Y	Y
Instruments				
ln HLA Het. × Post-1940	Y	Y	Y	Y
ln HLA Het. × Post-1980	N	N	N	Y
Observations	336	336	240	336
First Stage F Stat	87.74	16.93	25.75	12.73

Panel B: ln Labor Force				
ln Mortality Rate from Infectious Disease	-0.57***	-0.74**	-0.39**	-0.17
	(0.21)	(0.32)	(0.19)	(0.11)
Baseline Controls	N	Y	Y	Y
Instruments				
ln HLA Het. × Post-1940	Y	Y	Y	Y
ln HLA Het. × Post-1980	N	N	N	Y
Observations	336	336	240	336
First Stage F Stat	87.74	16.93	25.75	12.73

Panel C: ln Years of Schooling				
ln Mortality Rate from Infectious Disease	-0.19***	-0.47***	-0.33***	-0.22***
	(0.05)	(0.13)	(0.09)	(0.06)
Baseline Controls	N	Y	Y	Y
Instruments				
ln HLA Het. × Post-1940	Y	Y	Y	Y
ln HLA Het. × Post-1980	N	N	N	Y
Observations	336	336	240	336
First Stage F Stat	87.74	16.93	25.75	12.73

Panel E: ln Income per capita				
ln Mortality Rate from Infectious Disease	-0.41**	-0.28**	-0.16	-0.06
	(0.17)	(0.13)	(0.10)	(0.08)
Baseline Controls	N	Y	Y	Y
Instruments				
ln HLA Het. × Post-1940	Y	Y	Y	Y
ln HLA Het. × Post-1980	N	N	N	Y
Observations	336	336	240	336
First Stage F Stat	87.74	16.93	25.75	12.73

Panel A: ln Population				
ln Life Expectancy	7.62**	9.66**	7.78**	11.01**
	(2.90)	(3.61)	(2.92)	(4.38)
Baseline Controls	N	Y	Y	Y
Instruments				
ln HLA Het. × Post-1940	Y	Y	Y	Y
ln HLA Het. × Post-1980	N	N	N	Y
Observations	336	336	240	336
First Stage F Stat	205.23	59.84	44.67	43.56

Panel B: ln Labor Force				
ln Life Expectancy	7.54**	7.04*	5.42*	8.27*
	(2.97)	(3.57)	(2.90)	(4.32)
Baseline Controls	N	Y	Y	Y
Instruments				
ln HLA Het. × Post-1940	Y	Y	Y	Y
ln HLA Het. × Post-1980	N	N	N	Y
Observations	336	336	240	336
First Stage F Stat	205.23	59.84	44.67	43.56

Panel C: ln Years of Schooling				
ln Life Expectancy	2.46***	4.42***	4.60***	4.14***
	(0.58)	(0.71)	(0.80)	(0.65)
Baseline Controls	N	Y	Y	Y
Instruments				
ln HLA Het. × Post-1940	Y	Y	Y	Y
ln HLA Het. × Post-1980	N	N	N	Y
Observations	336	336	240	336
First Stage F Stat	205.23	59.84	44.67	43.56

Panel E: ln Income per capita				
ln Life Expectancy	5.37**	2.68**	2.20*	3.14***
	(2.40)	(1.12)	(1.31)	(1.09)
Baseline Controls	N	Y	Y	Y
Instruments				
ln HLA Het. × Post-1940	Y	Y	Y	Y
ln HLA Het. × Post-1980	N	N	N	Y
Observations	336	336	240	336
First Stage F Stat	205.23	59.84	44.67	43.56