

# Beyond Barker: Advantageous Early-Life Environments Cushion the Genetic Risk for Ischaemic Heart Disease

Samuel Baker (University of Bristol)

Pietro Biroli (University of Bologna)

Stephanie von Hinke (University of Bristol)

Hans van Kippersluis (Erasmus University Rotterdam)

IGSS, Boulder Colorado

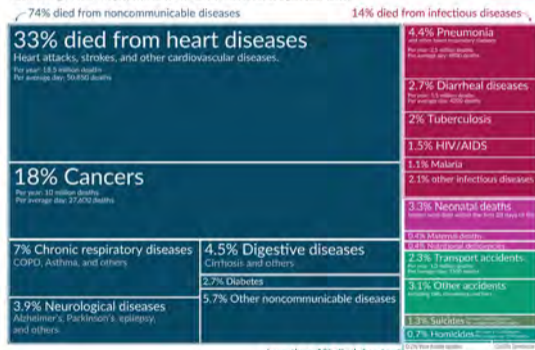
October 20th, 2023

# Cardiovascular Disease

## What do people die from? Causes of death globally in 2019

The size of the entire visualization represents the total number of deaths in 2019: 33 million. Each rectangle within it is proportional to the share of deaths due to a particular cause.

Our World in Data

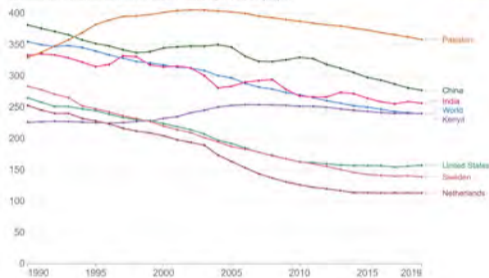


Data source: IHME Global Burden of Disease and Global Health Estimates. OurWorldInData.org - licensed under CC BY. All rights reserved. www.ourworldindata.org

## Cardiovascular disease death rate, 1990 to 2019

The estimated death rate from cardiovascular disease<sup>1</sup> per 100,000 people.

Our World in Data



Source: IHME Global Burden of Disease (2019). OurWorldInData.org/health-indicators/cv-death - CC BY. Note: To allow comparison between countries and over time this metric is age-standardized<sup>2</sup>.

**1. Cardiovascular disease:** Cardiovascular disease (CVD) refers to a range of conditions that affect the heart and blood vessels. They usually result from a buildup of fatty deposits in the arteries, which can restrict blood flow and lead to complications. This reduces the heart's ability to pump blood effectively to the heart muscle, brain, or other parts of the body, which can lead to consequences such as heart attacks, strokes, and heart failure.

**2. Age standardization:** Age standardization is an adjustment that makes it possible to compare populations with different age structures by standardizing them to a common reference population. <sup>3</sup> Read more: How does age standardization make health metrics comparable?

## The 'Fetal Origins' Hypothesis (Nurture)

- Adverse conditions *in utero* or very early  $\Rightarrow$  health and development
- Barker & Osmond (1986): *area-level* infant mortality 1920s vs. Ischaemic Heart Disease (IHD) mortality 1970s.

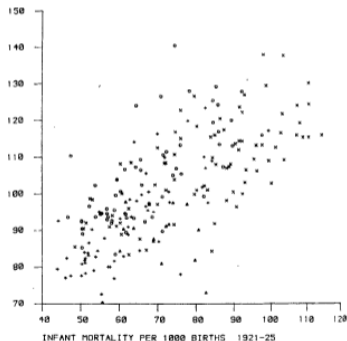


Fig 1—SMRs for ischaemic heart disease in 1968-78 at ages 35-74, men and infant mortality per 1000 births in 1921-25 in the 212 areas of England and Wales.

# This paper: GxE interplay

We study GxE interplay in ischaemic heart disease

- Does genetic susceptibility aggravate early-life circumstances?
- Can advantageous environments cushion genetic inequality?

# Contributions

- ① Replicate and extend Barker using high quality data, large sample sizes, *individual* level outcomes, and within-family design
  - Digitize UK infant mortality rates in Local Government Districts, and link with the UK Biobank (UKB)
  - $\sim 35$  years and  $\sim 1400$  districts of birth of the UKB participants
- ② Estimate causal effect of genetic predisposition on later life conditions

# Data

## UK Biobank:

- Drop non-European ancestry, those with missing values, and those born outside England & Wales
- Final sample size  $N = 378,838$  and  $N = 33,060$  full siblings

## Outcome variable: Ischaemic Heart Disease (IHD)

- Identified using the ICD-10 codes (I20-I25) obtained from mortality records and hospital inpatient records
- Around 11% of our sample has been diagnosed with IHD

## Independent variable:

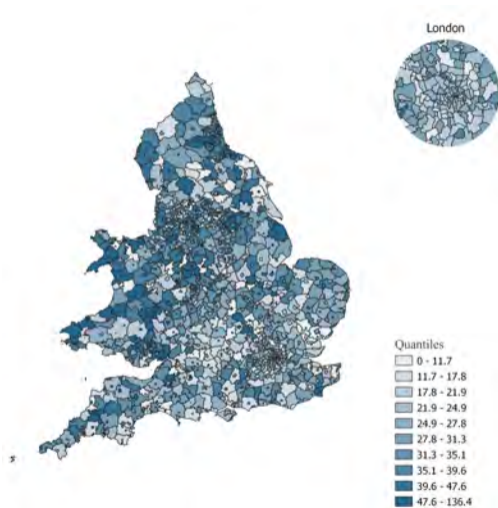
- Infant Mortality Rate (IMR)
  - Identify the Local Government District individuals were born in ( $n \approx 1400$ )
  - Digitize and merge the Infant Mortality Rate (number of deaths within the first year of life per 1,000 live births) in the district and year of birth
- Polygenic Index (PGI)
  - LDpred score using sumstats from Nikpay et al. (2015)

# Data – Infant Mortality Rate

TABLE 12.—continued.

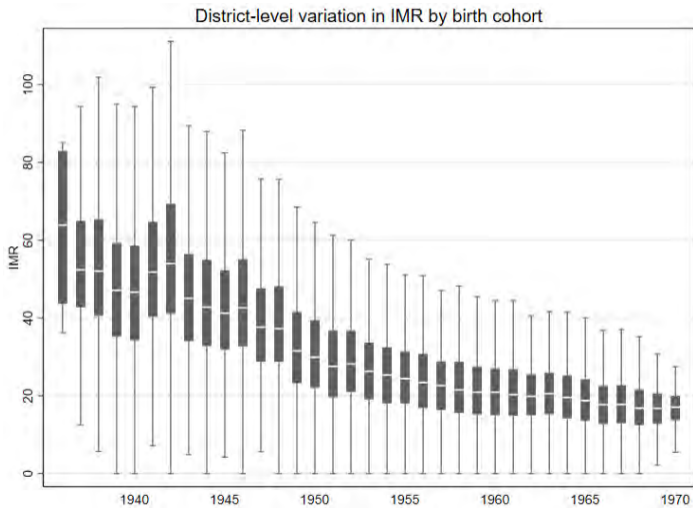
AREA	Estimated Home Population (Middle 1954)	LIVE BIRTHS								DEATHS						Deaths under 1 year		Deaths under 4 weeks	Summings	
		Legitimate		Illegitimate		TOTAL	Crude Rate per 1,000 Home Population	Comparability Factor	Ratio of local adjusted rate to National Rate	M.	F.	TOTAL	Crude Rate per 1,000 Population	Comparability Factor	Ratio of local adjusted death rate to National Rate	Number	Rate per 1,000 live births			
		M.	F.	M.	F.															(17)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	
<b>YORKSHIRE :</b>																				
<b>EAST RIDING...</b>	517,000	4,202	2,932	224	217	8,576	16.4	1.01	1.00	2,080	2,012	5,892	11.4	1.02	1.03	257	30	160	215	
County Borough :																				
Kilnspurn upon Hull...	206,000	2,683	2,492	168	160	5,503	18.3	0.96	1.06	1,713	1,492	3,205	10.7	1.15	1.09	106	34	306	340	
Admin. County...	217,000	1,520	1,440	54	57	3,872	14.2	1.09	1.02	1,367	1,320	2,687	12.4	0.88	0.97	71	23	94	64	
Urban Districts...	106,600	746	683	27	33	1,480	14.0	1.07	0.99	721	760	1,481	13.9	0.83	1.02	34	23	25	25	
Beverley M.B.	15,300	117	120	5	7	230	17.5	1.00	1.15	95	94	189	12.3	0.91	0.99	5	5	8	4	
Bridlington M.B.	24,470	161	135	10	10	313	12.8	1.04	0.96	191	220	420	17.1	0.71	1.08	1	9	20	4	
Driffield	6,900	41	33	2	3	79	11.4	0.97	0.73	39	42	81	11.7	0.75	0.78	1	1	1	1	
Fife	4,790	28	20	—	1	57	11.9	1.06	0.83	32	29	61	12.7	0.86	0.97	1	1	1	1	
Halewippen	37,720	254	250	6	6	516	13.7	1.08	0.97	251	240	491	13.0	0.85	1.00	10	19	5	12	
Hedon M.B.	2,090	18	18	1	—	37	17.7	1.04	1.21	14	16	30	14.4	0.89	1.13	—	27	—	—	
Hornsea	5,450	41	32	1	—	74	13.6	1.04	1.02	40	50	90	16.5	0.67	0.98	—	27	—	—	
Norton	4,330	28	32	3	1	74	15.3	1.09	1.11	25	29	54	11.2	0.83	0.92	—	27	—	—	
Withernsea	4,960	28	35	1	1	69	13.9	1.08	1.01	34	31	65	13.1	0.77	0.89	—	26	—	—	
Rural Districts...	110,400	774	757	29	24	1,584	14.3	1.00	1.03	646	560	1,206	10.9	0.93	0.90	37	23	20	41	
Beverly	21,620	159	145	5	2	311	14.4	1.13	1.07	137	111	248	11.5	1.02	1.04	8	26	6	8	
Bridlington	9,450	56	61	4	7	128	13.5	1.13	1.00	55	50	105	11.1	0.88	0.86	3	25	1	1	
Derwent	13,940	90	86	—	1	177	12.7	1.03	0.86	77	62	139	10.0	0.94	0.83	6	34	—	8	
Driffield	11,270	92	79	3	3	168	14.9	1.14	1.12	74	53	127	11.3	0.94	0.94	4	24	—	5	
Huddersfield	19,470	123	137	4	5	272	13.7	1.13	1.02	120	118	238	12.0	0.90	0.96	7	26	—	7	
Howden	11,680	73	90	6	6	175	15.0	1.08	1.07	56	66	122	10.4	0.88	0.81	—	23	—	5	
Norton	7,530	58	59	—	—	117	15.5	1.08	1.10	34	35	69	9.2	0.84	0.77	—	17	—	1	
Pocklington	15,000	118	109	7	2	236	15.7	1.08	1.12	93	65	158	10.5	0.92	0.85	—	13	—	4	
<b>YORKSHIRE :</b>																				
<b>NORTH RIDING...</b>	529,600	4,554	4,187	214	205	9,164	17.3	1.04	1.08	3,159	2,854	6,015	11.4	1.06	1.07	272	30	181	225	
County Borough :																				
Middlesbrough	548,100	1,468	1,431	75	85	3,059	20.7	0.94	1.28	916	727	1,643	11.1	1.24	1.22	104	34	68	77	

# Data – descriptives IMR

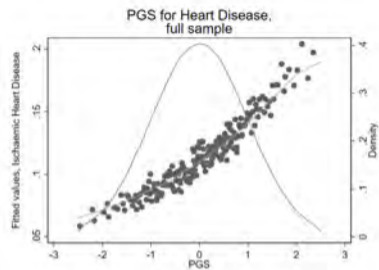
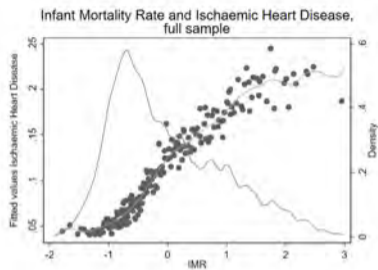




# Data – descriptives IMR



# Data – descriptives IMR & PGI



# Empirical specification

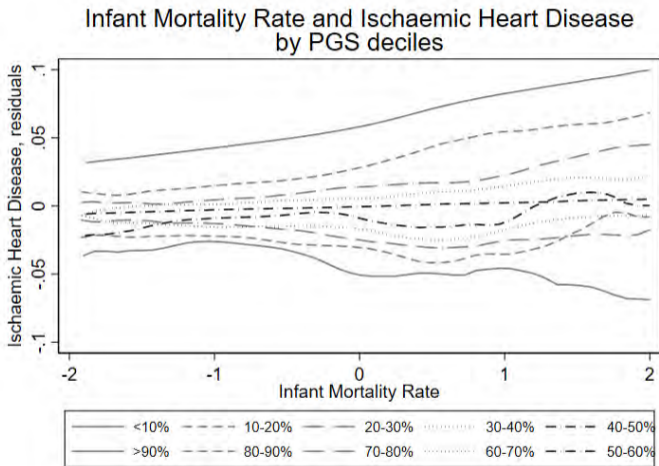
$$Y_{ijd} = \alpha + \beta_1 IMR_{d,t=0} \times PGI_i + \beta_2 IMR_{d,t=0} + \beta_3 PGI_i + \gamma \mathbf{X}_i + u_{id} \quad (1)$$

$$Y_{ijd} = \tilde{\alpha} + \tilde{\beta}_1 IMR_{d,t=0} \times PGI_i + \tilde{\beta}_2 IMR_{d,t=0} + \tilde{\beta}_3 PGI_i + \tilde{\gamma} \mathbf{X}_i + \delta_d + \tilde{u}_{id} \quad (2)$$

$$Y_{ijd} = \dot{\alpha} + \dot{\beta}_1 IMR_{d,t=0} \times PGI_i + \dot{\beta}_2 IMR_{d,t=0} + \dot{\beta}_3 PGI_i + \dot{\gamma} \mathbf{X}_i + \eta_j + \dot{u}_{id} \quad (3)$$

- $Y_{id}$  is IHD of individual  $i$ , in family  $j$ , born in district  $d$
- $IMR_{d,t=0}$  and  $PGI_i$  standardized (mean 0, std 1)
- $\mathbf{X}_i$  are individual characteristics
  - Sex, YoB  $\times$  MoB dummies, 10 principal components, square terms of  $IMR_{d,t=0}$  and  $PGI_i$ , and interactions between  $\mathbf{X}_i$  and  $IMR_{d,t=0}$  and  $PGI_i$
- $u_{id}$  is the error term, clustered by district
- $\delta_d$  are district indicators / fixed-effects
- $\eta_j$  are family indicators / fixed-effects

# Raw Data



# OLS: main regression

	Ischaemic Heart Disease				
	(1)	(2)	(3)	(4)	(5)
IMR	0.0148*** (0.0016)		0.0120*** (0.0013)	0.0122*** (0.0013)	0.0038*** (0.0010)
IMR <sup>2</sup>	-0.0008 (0.0005)		-0.0004 (0.0005)	-0.0005 (0.0005)	0.0005 (0.0004)
PGI		0.0291*** (0.0006)	0.0289*** (0.0006)	0.0289*** (0.0006)	0.0285*** (0.0006)
PGI <sup>2</sup>		0.0042*** (0.0004)	0.0042*** (0.0004)	0.0042*** (0.0004)	0.0042*** (0.0004)
IMR × PGI				0.0097*** (0.0006)	0.0098*** (0.0006)
Covariates	Yes	Yes	Yes	Yes	Yes
District FEs	No	No	No	No	Yes
Mean	0.11	0.11	0.11	0.11	0.11
R <sup>2</sup>	0.07	0.07	0.07	0.08	0.08
No. of observations	378'807	378'807	378'807	378'807	378'807

Notes: 'Mean' is the mean of the dependent variable. Robust standard errors clustered by district in parentheses. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

## Results - Family fixed effects

	Reduced (sibling) sample			With family fixed effects		
IMR	0.0078** (0.0036)	0.0068* (0.0036)	0.0071** (0.0036)	0.0012 (0.0058)	0.0010 (0.0058)	0.0014 (0.0058)
IMR <sup>2</sup>	0.0016 (0.0016)	0.0019 (0.0016)	0.0019 (0.0016)	0.0062*** (0.0022)	0.0064*** (0.0022)	0.0064*** (0.0022)
PGI		0.0276*** (0.0017)	0.0273*** (0.0017)		0.0223*** (0.0033)	0.0220*** (0.0033)
PGI <sup>2</sup>		0.0059*** (0.0013)	0.0059*** (0.0012)		0.0057*** (0.0019)	0.0057*** (0.0019)
IMR × PGI			0.0049*** (0.0018)			0.0053** (0.0026)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
District FE	No	No	No	No	No	No
R <sup>2</sup>	0.07	0.07	0.07	0.56	0.57	0.57
N	33'058	33'058	33'058	33'058	33'058	33'058

Notes: Columns (1)-(3) show robust standard errors clustered by district on the full UKB sample. Columns (4)-(6) show robust standard errors clustered by family on the UKB sibling sample. Columns (7)-(9) show robust standard errors clustered by family on the sibling sample. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

## Summary of results: Environment

Barker hypothesis:

- Adverse circumstances (e.g., high IMR) in early life associated with the prevalence of IHD later in life
    - Association works at area *and* individual level
    - Attenuates if looking at within-district variation
    - Almost zero within families
- ⇒ IMR: between families unobservables (e.g., poverty, health behavior, etc.)

## Summary of results: Genes

- Genetic predispositions play an important role
  - Levels  $G$ : main effect 'bigger' than IMR
  - Variation  $G \times E$ : amplified by the environment
- Causal effect
  - Within-family variation in PGI is quasi-random
  - Inequality at birth close to zero in advantageous conditions



# Limitations

- UK Biobank is not representative
- There will be selective mortality (by definition!)
- Estimates of the IMR cannot be interpreted causally
  - Cannot identify the exact mechanism
  - But likely to capture predetermined “environmental influences” as it is uncorrelated with genetic predisposition, at least within-families

Hence, environmental interventions may reduce genetic susceptibility to heart disease

Thank you

# Families in the UK Biobank

Full siblings can be identified using their genotypes

- They share 50% of their genetic code

Approximately 35,000 siblings in the full data

- Drop ~5,000 due to missing data, not born in England or Wales
- Big reduction in the sample size
- Exploiting variation *within families*
- But there is relatively little *within-family* variation
- Loss of power

	CHD	IMR	PGI CHD
Overall SD	0.232	19.1	0.995
Between SD	0.169	16.3	0.858
Within SD	0.160	10.3	0.507

## Results - by gender

	Females					Males				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
IMR	0.0088*** (0.0014)		0.0088*** (0.0014)	0.0089*** (0.0014)	0.0008 (0.0012)	0.0159*** (0.0017)		0.0159*** (0.0017)	0.0161*** (0.0017)	0.0071*** (0.0017)
IMR <sup>2</sup>	0.0007 (0.0005)		0.0007 (0.0005)	0.0007 (0.0005)	0.0017*** (0.0005)	-0.0014** (0.0007)		-0.0014** (0.0007)	-0.0014** (0.0007)	-0.0003 (0.0006)
PGI		0.0169*** (0.0005)	0.0170*** (0.0005)	0.0171*** (0.0005)	0.0169*** (0.0005)		0.0429*** (0.0010)	0.0429*** (0.0010)	0.0424*** (0.0010)	0.0420*** (0.0010)
PGI <sup>2</sup>		0.0028*** (0.0004)	0.0028*** (0.0004)	0.0028*** (0.0004)	0.0028*** (0.0004)		0.0058*** (0.0007)	0.0058*** (0.0007)	0.0059*** (0.0006)	0.0059*** (0.0007)
IMR × PGI				0.0068*** (0.0007)	0.0069*** (0.0007)				0.0129*** (0.0010)	0.0130*** (0.0010)
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FEs	No	No	No	No	Yes	No	No	No	No	Yes
Mean	0.07	0.07	0.07	0.07	0.07	0.16	0.16	0.16	0.16	0.16
R <sup>2</sup>	0.04	0.04	0.04	0.04	0.05	0.07	0.07	0.07	0.07	0.08
No. of observations	203'732	203'732	203'732	203'732	203'720	175'076	175'076	175'076	175'076	175'055

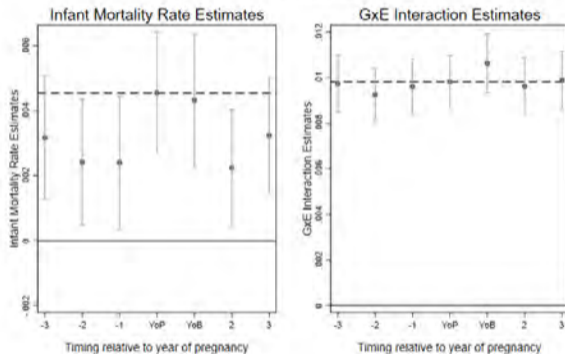
Notes: 'Mean' is the mean of the dependent variable. Robust standard errors clustered by district in parentheses. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

## Further robustness checks

	Ischaemic Heart Disease				
	(1)	(2)	(3)	(4)	(5)
IMR	0.0014 (0.0058)	0.0014 (0.0060)	0.0015 (0.0058)	0.0014 (0.0058)	0.0093* (0.0049)
IMR <sup>2</sup>	0.0064*** (0.0022)	0.0065*** (0.0022)	0.0062*** (0.0022)	0.0068*** (0.0025)	
PGI	0.0220*** (0.0033)	0.0220*** (0.0033)	0.0323*** (0.0032)	0.0220*** (0.0033)	0.0216*** (0.0033)
PGI <sup>2</sup>	0.0057*** (0.0019)	0.0057*** (0.0019)	0.0054*** (0.0019)	0.0061*** (0.0022)	
IMR × PGI	0.0053** (0.0026)	0.0053** (0.0026)	0.0088*** (0.0029)	0.0053** (0.0026)	0.0054** (0.0026)
IMR <sup>2</sup> × PGI <sup>2</sup>				-0.0004 (0.0011)	
Illegitimacy Rate		-0.0005 (0.0045)			
Birth Rate		0.0022 (0.0042)			
Death Rate		-0.0013 (0.0041)			
Covariates	Yes	Yes	Yes	Yes	Yes

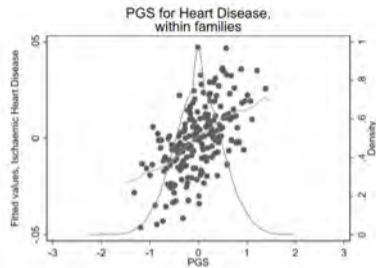
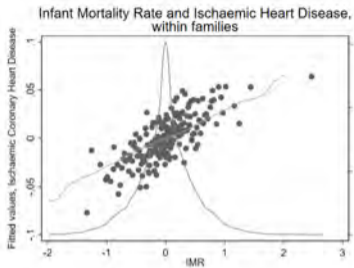
# Timing of the Infant Mortality Rate

Timing of IMR measurement relative to year of pregnancy with district fixed effects



**Figure:** Exploring the timing of the infant mortality rate for Ischaemic Heart Disease; full sample, controlling for district fixed effects.

# Data – descriptives IMR & PGI



## IMR variation

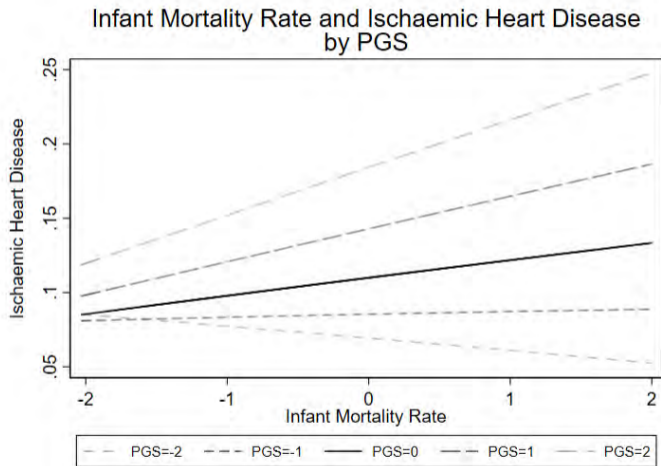
- Rely on variation in IMR across districts within cohorts
- Areas with higher IMR also poorer + worse access to medical care

Table: Correlations between district level IMR with social class

	All years	1951	1961	1971
Illegitimacy rate	0.0108 (0.0087)	0.0050 (0.0238)	-0.0045 (0.0145)	0.0286*** (0.0095)
Proportion social class I	-0.3391*** (0.1229)	-0.7338*** (0.2652)	-0.4186** (0.2100)	-0.2260 (0.1506)
Proportion social class II	-0.0309 (0.0374)	0.1142 (0.0875)	-0.0637 (0.0498)	-0.0851 (0.0626)
Proportion social class IV	0.0990** (0.0419)	0.1791*** (0.0650)	0.0581 (0.0693)	-0.1434** (0.0719)
Proportion social class V	0.2417*** (0.0745)	0.3707*** (0.1138)	0.2284 (0.1475)	0.1034 (0.1350)
R <sup>2</sup>	0.17	0.03	0.03	0.03
No. of observations	4'349	1'470	1'411	1'468



## OLS: visuals



# Health Behaviors

What could explain this relationship?

- Consider adult intermediary behaviors and outcomes:
  - Body Mass Index
  - Height
  - Blood Pressure
  - Smoking
  - Drinking

# Potential mediators in the gene-environment interplay for ischaemic heart disease

	BMI		DBP		SBP		Height		Alcohol		Smoking	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
IMR	0.2476*** (0.0570)	-0.0108 (0.0682)	0.1825 (0.1817)	0.0549 (0.1675)	0.4441 (0.2974)	0.3351 (0.2908)	-0.7331*** (0.0888)	0.0865 (0.0821)	-0.0004 (0.0021)	0.0001 (0.0030)	-0.0075 (0.0058)	-0.0084 (0.0080)
IMR <sup>2</sup>	-0.0480** (0.0188)	0.0175 (0.0278)	-0.0213 (0.0548)	-0.0049 (0.0649)	0.0497 (0.0851)	0.0362 (0.1088)	0.1208*** (0.0321)	-0.0139 (0.0288)	-0.0006 (0.0008)	-0.0009 (0.0012)	0.0027 (0.0023)	0.0030 (0.0032)
PGI	0.3108*** (0.0277)	0.2769*** (0.0401)	0.5133*** (0.0526)	0.4637*** (0.0936)	1.0752*** (0.1118)	0.9571*** (0.1857)	-0.3922*** (0.0385)	-0.1045** (0.0435)	-0.0032*** (0.0010)	-0.0042** (0.0020)	0.0140*** (0.0029)	0.0104** (0.0048)
PGI <sup>2</sup>	-0.0068 (0.0186)	-0.0040 (0.0243)	-0.0786* (0.0423)	-0.0581 (0.0573)	-0.1027 (0.0785)	-0.0582 (0.1128)	0.0380 (0.0271)	0.0009 (0.0292)	-0.0009 (0.0008)	-0.0002 (0.0011)	-0.0012 (0.0019)	-0.0026 (0.0026)
IMR × PGI	-0.0537* (0.0276)	-0.0268 (0.0358)	-0.2010*** (0.0553)	-0.1991** (0.0822)	-0.0570 (0.1102)	-0.0138 (0.1473)	-0.0372 (0.0352)	0.0003 (0.0418)	0.0001 (0.0012)	-0.0024 (0.0018)	-0.0045 (0.0031)	-0.0060 (0.0043)
Covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Family FEs	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Mean	27.29	27.29	82.10	82.10	137.98	137.98	168.11	168.11	0.96	0.96	0.45	0.45
R <sup>2</sup>	0.03	0.64	0.06	0.60	0.13	0.63	0.54	0.89	0.03	0.56	0.03	0.60
No. of observations	33'276	33'276	31'840	31'840	31'840	31'840	31'840	31'840	31'797	31'797	31'592	31'592

Notes: BMI is measured as weight (in kilograms) divided by height (in metres squared). DBP and SBP denote diastolic and systolic blood pressure, respectively, and are measured in mmHg (millimeters of mercury). Height is measured in cm. Alcohol and smoking are dummy variables indicating whether the individual currently smokes or drinks. 'Mean' is the mean of the dependent variable. Robust standard errors clustered by district (column 1) and family and district (column 2) in parentheses. \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .