Online Appendix

“The Roots of Health Inequality and The Value of Intra-Family Expertise”

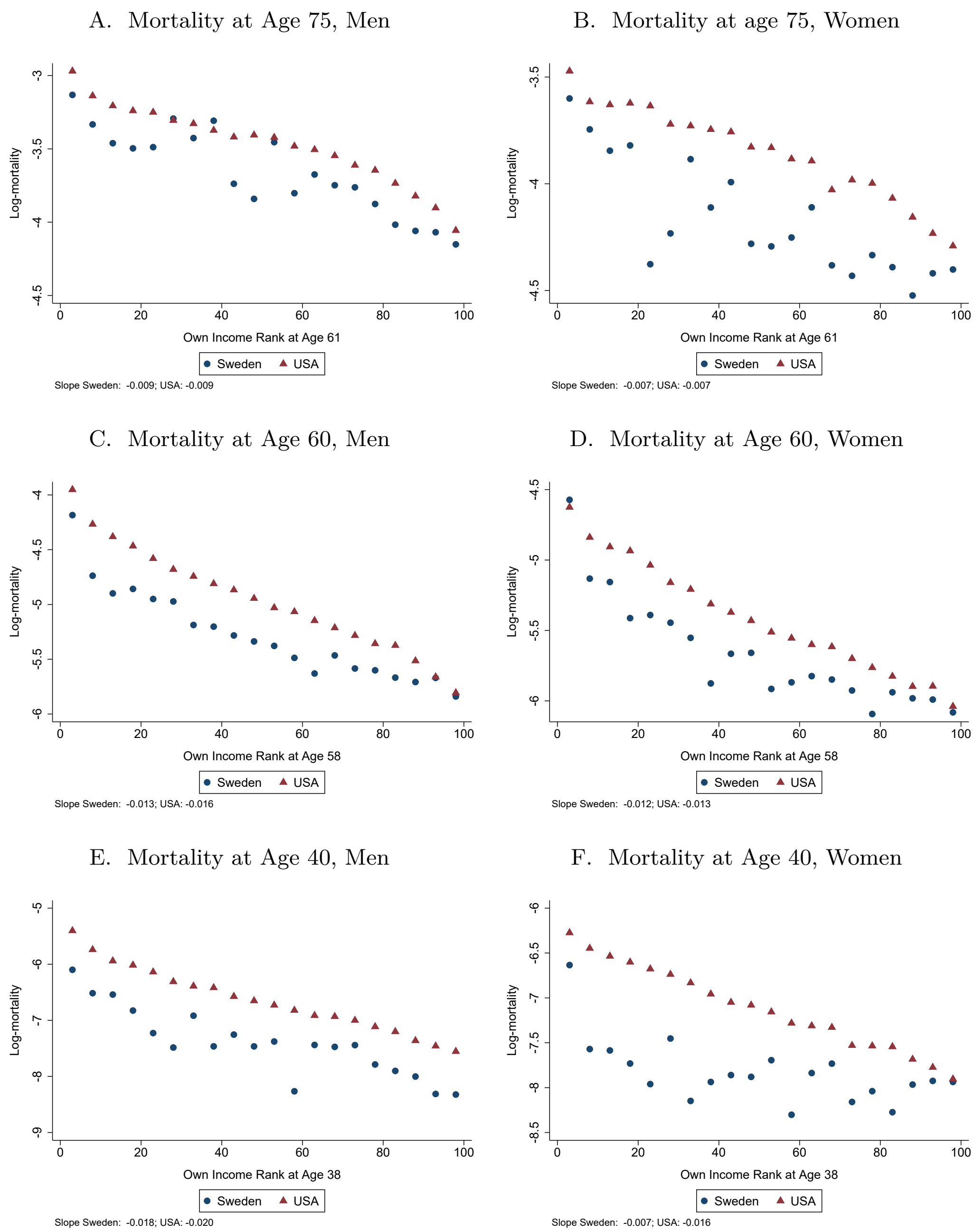
Yiqun Chen

Petra Persson

Maria Polyakova

Appendix figures and tables

Figure A1: Income Gradients in Mortality in the US and Sweden

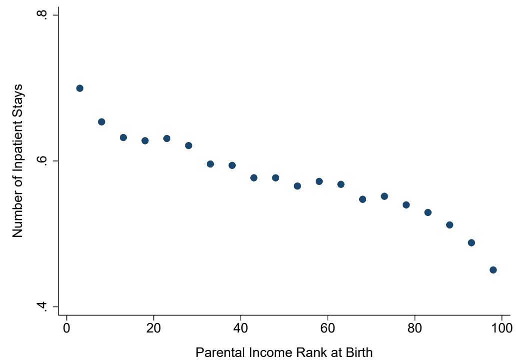


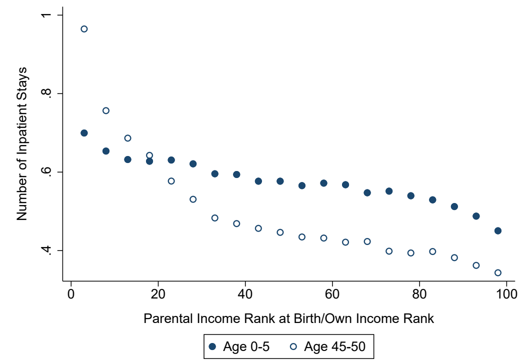
Figures compare mortality rates by income rank ventile between Sweden and the US. We plot log-mortality at age 40, 60 and 75 conditional on ventiles of income rank at age 38, 58, and 60 (Sweden) or 61 (US), respectively. The latter measures incomes a year before the earliest retirement ages in the respective countries. US mortality data is derived from the data reported by The Health Inequality Project. The sample for Sweden is restricted to the same birth cohorts as in the US data. Income is measured as the equivalent of the US adjustable gross income (includes work-related income, self-employment income, and capital income). We exclude individuals with zero or negative income from the analysis. The note in each panel reports the estimated slope of a linear regression of log-mortality rate on income rank percentile, separately by country and gender. We cannot reject the statistical equivalency of the slopes measuring the mortality gradient for men at age 40 and at age 75 (p-values 0.46 and 0.87, respectively), as well as for women at age 75 (p-value 0.96). We reject the equivalency for the mortality gradient for men at age 60 and for women at age 40 (p-value 0.00), and for women at age 60 at the 10% significance level (p-value 0.05).

Figure A2: Early Emergence of the Health-SES Gradient

A. Number of Inpatient Stays, by Age 5

B. Number of Inpatient Stays over Life Cycle





Figures plot the share of individuals with relevant outcomes for each ventile of income rank. Parental income rank at birth is assigned based on the average of parental incomes in the two years before the child was born relative to other parents with children in the same birth cohort. Income rank for adults aged 45-50 in Panel B are assigned based on each individual's own income at age 40 relative to other individuals in the same gender-birth cohort. Inpatient stays due to pregnancy, childbirth and the puerperium are excluded from the count of inpatient stays in Panels A and B.

Figure A3: Share of population with a doctor or a nurse in the family

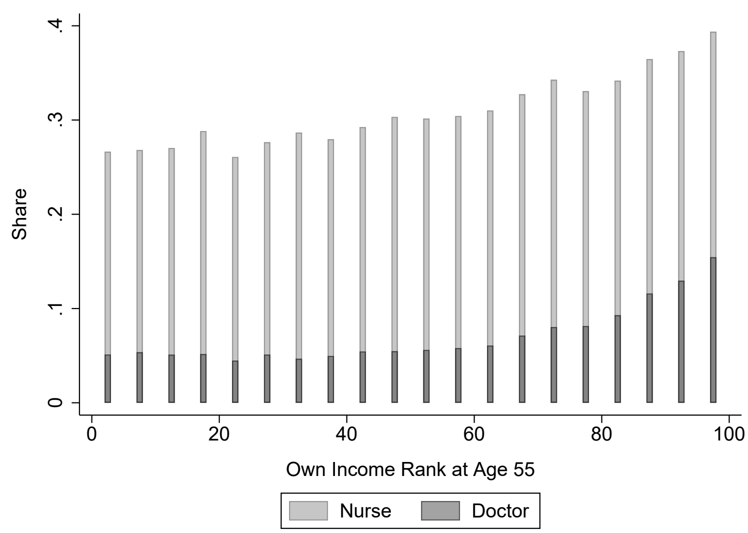


Figure plots the share of individuals with a nurse or a doctor in the family by income ranks. The sample of individuals is defined as in Panel A of Figure 2.

Figure A4: Tobacco Exposure, *in utero*

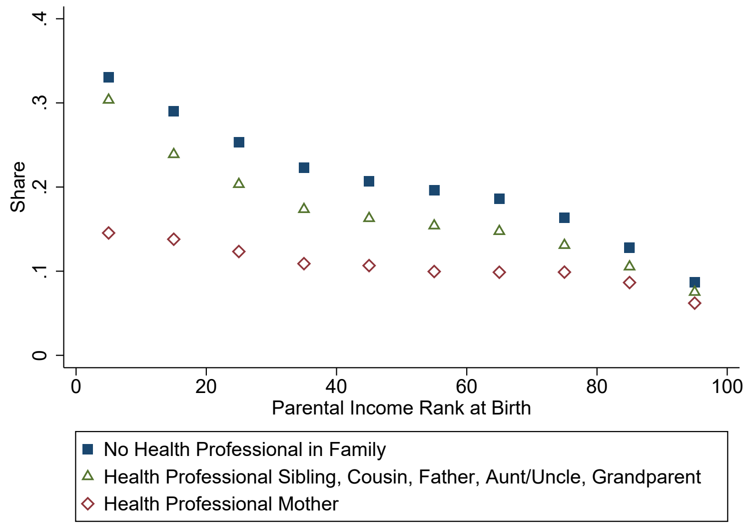
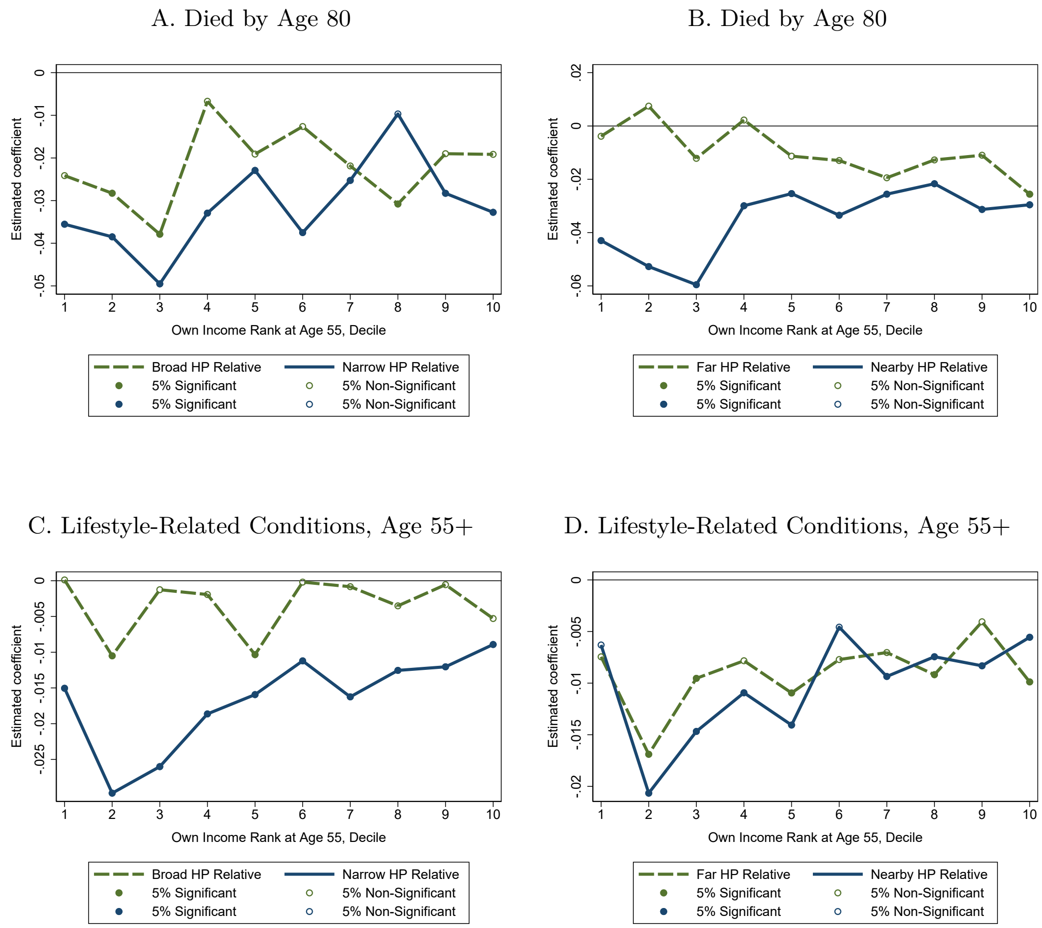


Figure plots the share of children exposed to tobacco *in utero* by parental income rank at birth. Parental income rank at birth is assigned based on the average of parental incomes in the two years before the child was born relative to other parents with children in the same birth cohort. We start with the same data sample as defined in Panel C of Figure 3. The sample is split by whether an individual has a health professional relative or a health professional mother. Individuals are assigned to the sample "with a health professional" if at least one member of their broad family (sibling, cousin, father, aunt/uncle, grandparent) has a university degree in medicine or nursing. Individuals are assigned to the sample "with a health professional mother" if the mother has a university degree in medicine or nursing.

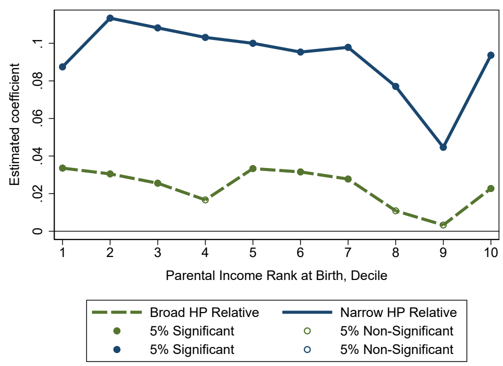
Figure A5: Health Professional in the Family and Health at Older Ages: Heterogeneity



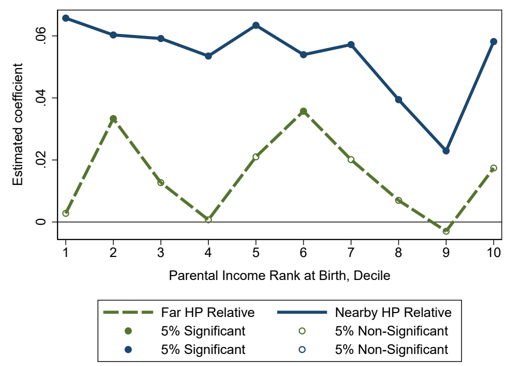
Figures replicate the analyses in Panels B and D (specifications with the full set of covariates) of Figure 2 for sub-samples of the data. Panels A and C re-estimate conditional differences in mortality and the prevalence of lifestyle-related conditions separately for individuals that have a broad, but no narrow, health professional in the family (dashed line), and for individuals that have a narrow health professional (solid line), relative to individuals that have no health professionals. A broad family tie is defined as having a health professional, who is a sibling, cousin, niece/nephew, or a grandchild. A narrow family tie is defined as having a health professional, who is a child, child-in-law, or a spouse. Panels B and D split the same based on geographic proximity. An individual is defined to have a nearby health professional relative (solid line) if both reside in the same county in the same year for more than 50 percent of the years that are observed between 1991 and 2016. The individual is defined to have a far health professional relative otherwise (dashed line). In all regressions, the comparison group is the set of individuals without any health professional relative. Coefficients are reported as filled circles if they are estimated with at least 5% statistical significance level, and as hollow circles otherwise.

Figure A6: Health Professional in the Family and Health at Younger Ages: Heterogeneity

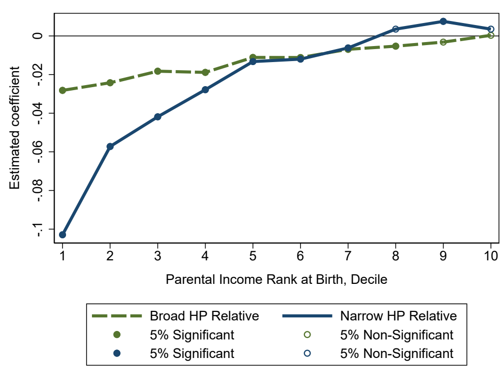
A. HPV Vaccine, by Age 20



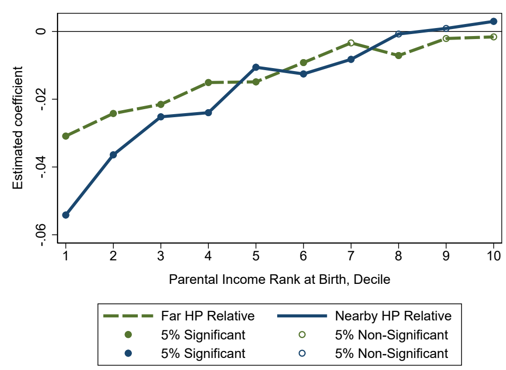
B. HPV Vaccine, by Age 20



C. Tobacco Exposure, *in utero*



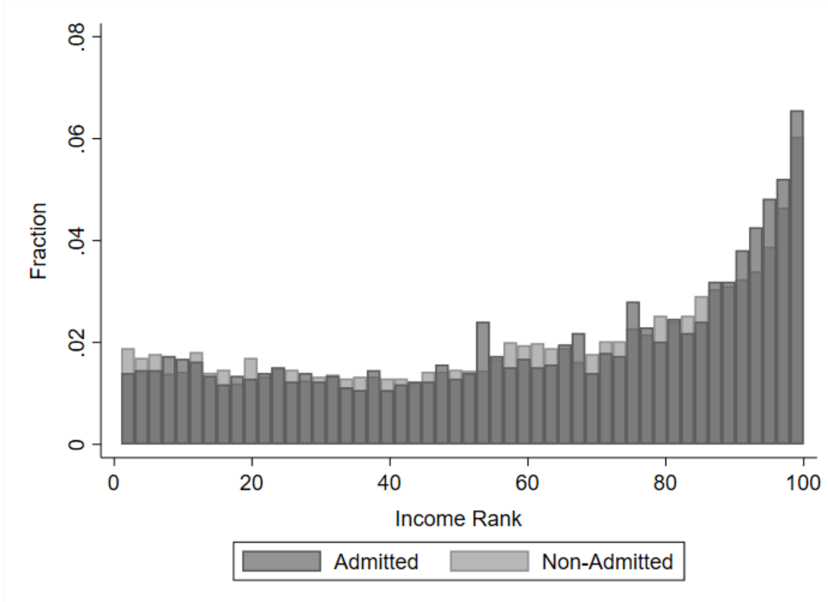
D. Tobacco Exposure, *in utero*



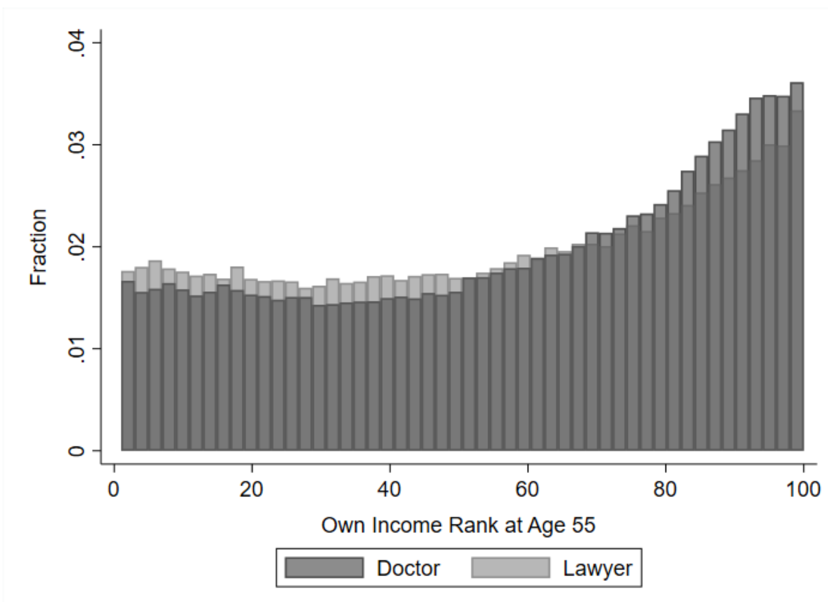
Figures replicate the analyses in Panels B and D (specifications with the full set of covariates) of Figure 3 for sub-samples of the data. Panels A and C re-estimate conditional differences in HPV vaccination and tobacco exposure *in utero* separately for individuals that have a broad, but no narrow, health professional in the family (dashed line), and for individuals that have a narrow health professional (solid line), relative to individuals that have no health professionals. A broad family tie is defined as having a health professional, who is a sibling, cousin, aunt/uncle, or a grandparent. A narrow family tie is defined as having a health professional mother or father. Panels B and D split the same based on geographic proximity. An individual in Panel B is defined to have a nearby health professional relative (solid line) if the individual's parents reside in the same county in the same year for more than 50 percent of the years that are observed between 1991 and 2016. A child in Panel D is defined to have a nearby health professional relative (solid line) if in the year of birth, a health professional relative lived in the same county as the mother. The individuals are defined to have a far health professional relative otherwise (dashed line). In all regressions, the comparison group is the set of individuals without any health professional relative. Coefficients are reported as filled circles if they are estimated with at least 5% statistical significance level, and as hollow circles otherwise.

Figure A7: Income Distribution of the Medical School Lottery and the Event Study Analysis Sample

A. Medical School Lottery Sample

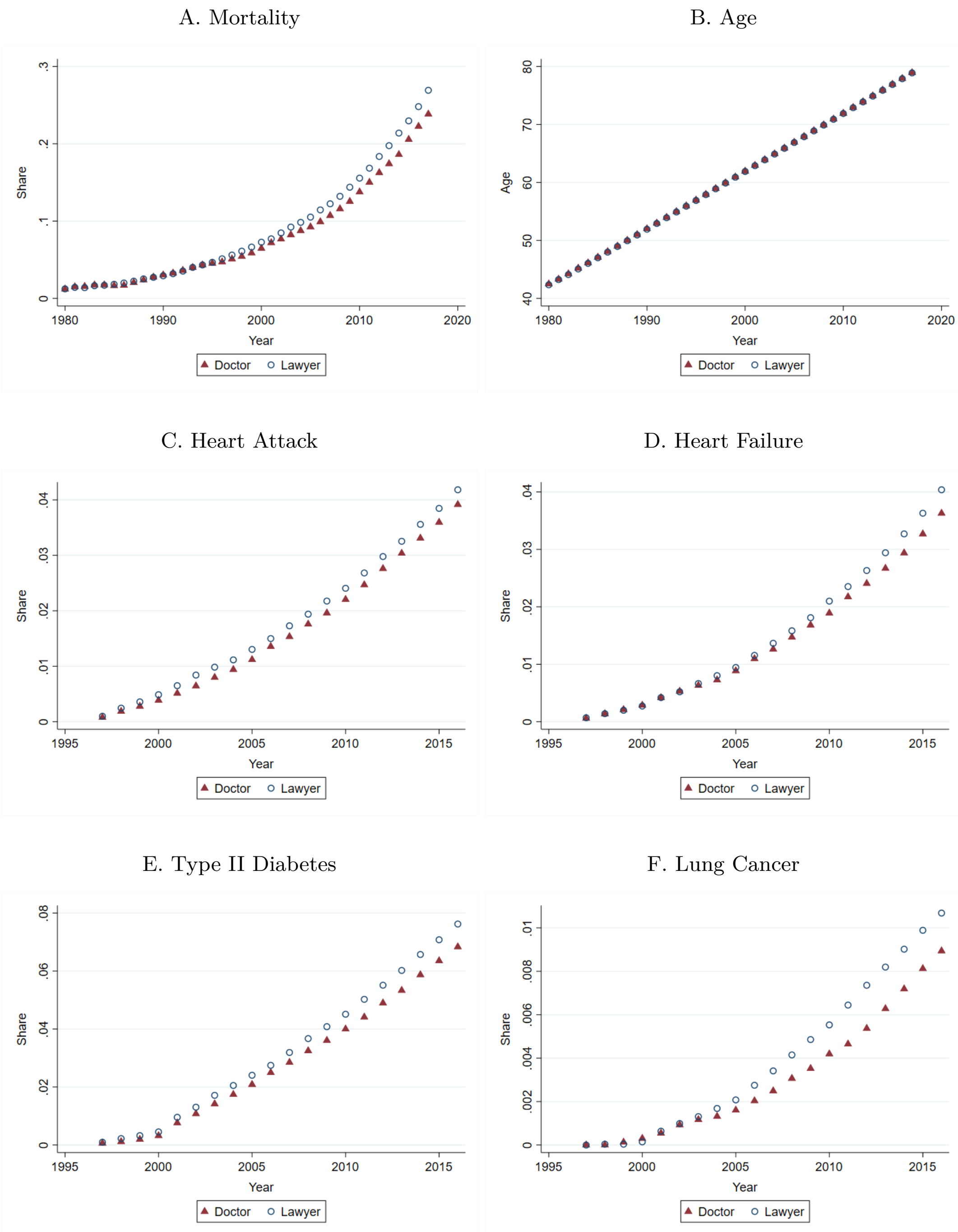


B. Event Study Sample



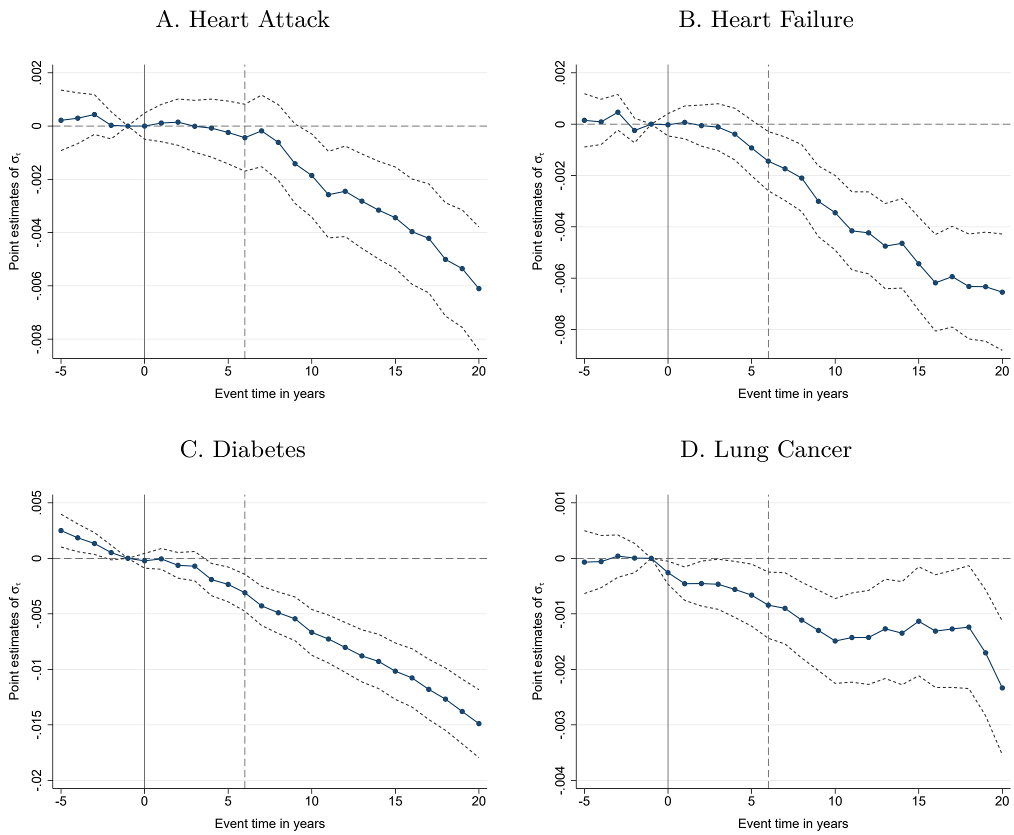
Panel A plots own income rank at age 30, 40, or 55 for the sample used in the medical school lottery analyses, individuals that are too young or too old to be observed income are not included in Panel A. Panel B plots own income rank at age 55 for the sample used in the lifestyle-related conditions index event study analyses (i.e., Panel B of Figure 5).

Figure A8: Doctor in the Family and Long-Run Health Bonus: Descriptive Evidence



Panel A records the cumulative share (y-axis) of individuals born in Sweden in 1936-1940, who have died by a given calendar year (x-axis). Panel B records the average age of the same individuals by calendar year, keeping deceased individuals in the sample. Panels C to F record the share of individuals born in Sweden between 1936 and 1961, who have acquired the specified chronic condition by a given calendar year (x-axis). Deceased individuals are kept in the balanced sample. In all panels the sample is restricted to individuals who at some point in their lifetime had a child matriculated in the study of either law or medicine. The outcomes are shown separately for the group of individuals whose child matriculated into medical (filled triangles) or legal (hollow circles) studies. We exclude observations if at least one parent is a health professional (a physician or a nurse) herself or himself. Parents with a child who became a nurse before another child became a doctor are not included in the "doctor" sample; parents that have a child trained as a lawyer and another child trained as a doctor (or nurse) are excluded from the "lawyer" sample.

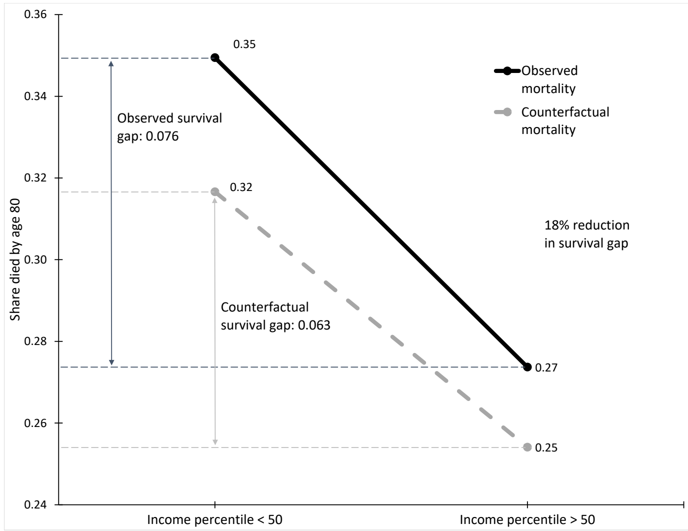
Figure A9: Doctor in the Family and Long-Run Health Bonus: Event Studies



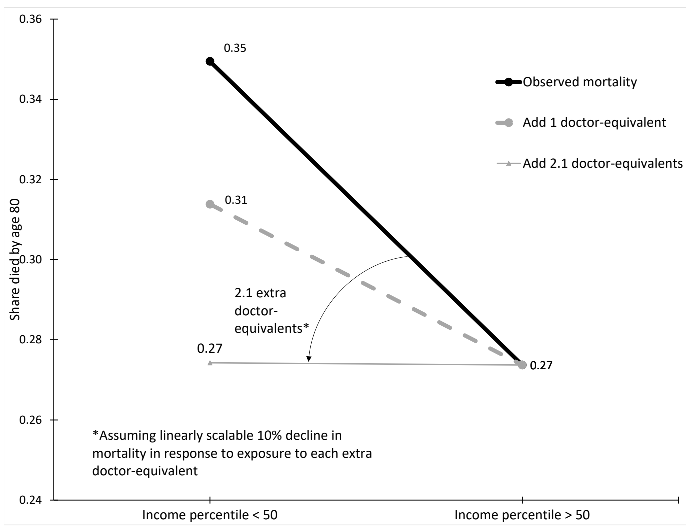
Figures plot coefficients σ_τ and 95% confidence intervals from the event study specification in Equation (4). The sample is restricted to family members of doctors and lawyers born in Sweden between 1936 and 1961. In both Panels, we exclude family members who are themselves a health professional, or have a health professional spouse. Family members with a relative became a nurse before another relative became a doctor are dropped from the "doctor" sample; family members with both a lawyer and a health professional relative are dropped from the "lawyer" sample. All panels exclude individuals that have died before the first year of clinical records (i.e., 1997). The regressions are centered at event year -1, i.e., one year before the year of matriculation in a medical or legal degree. The dashed vertical line marks the average graduation time for physicians. Standard errors are clustered at the family level.

Figure A10: Exposure to Expertise and Income-Mortality Gradient

A. Universal access to expertise



B. Health-income gradient in "doctor-equivalents"



Both panels plot observed and counterfactual gradients in mortality. The sample is defined as in Panel A of Figure 1. In both panels, the solid line ("observed mortality") plots observed average probability of individuals surviving until age 80 conditional on being alive at age 55, across the first half and the second half of the income rank distribution at age 55 (among 1936-1937 cohorts). In Panel A, the dashed line ("counterfactual mortality") is computed as follows. We multiply observed mortality within each income group by the treatment effect (1-T) and then re-scale the result by (1-T*s) to account for the differences in the underlying prevalence of access to expertise. T is the treatment effect of access to expertise (estimated at 10 percent for mortality, following the descriptive results in Appendix Table A1 and the event study estimates in Panel A of Figure 5); **s** is the share of individuals with health literacy at the baseline, proxied by the college completion rate (among 1936-1937 cohorts) of 7 percent and 31 percent percent at the bottom half and top half of the income distribution, respectively. The resulting formula for counterfactual mortality is: observed mortality within the income group times the treatment effect (1-0.1), divided by (1-0.1*0.07) for the first half of the income distribution, and divided by (1-0.1*0.31) for the second half. The denominator term re-scales the numerator to account for the differences in the baseline levels of access to expertise and hence the share of individuals that gets treated. In Panel B, the points on the dashed and light grey line are computed by applying 10 percent estimate of mortality reduction for exposure to each additional doctor at the bottom half of the income distribution.

Table A1: Health Professional in the Family and Health Outcomes

Panel A: Died, by Age 80

	(1) Income Decile 1	(2) Income Decile 2	(3) Income Decile 3	(4) Income Decile 4	(5) Income Decile 5	(6) Income Decile 6	(7) Income Decile 7	(8) Income Decile 8	(9) Income Decile 9	(10) Income Decile 10
Health Professional in Family	-0.059***	-0.058***	-0.073***	-0.039***	-0.043***	-0.045***	-0.045***	-0.026***	-0.035***	-0.036***
(no covariate)	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)	(0.007)
Health Professional in Family	-0.030***	-0.034***	-0.044***	-0.019**	-0.021**	-0.027***	-0.023***	-0.018**	-0.025***	-0.028***
(full covariates)	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)
Mean of Dep. Var.	0.41	0.36	0.34	0.31	0.30	0.30	0.29	0.27	0.26	0.23
Std. Dev. of Dep. Var.	0.49	0.48	0.47	0.46	0.46	0.46	0.46	0.44	0.44	0.42
Number of Observations	11,454	12,850	12,777	12,928	12,882	12,823	12,786	12,681	12,252	13,396

Panel B: Index of Lifestyle-Related Conditions, Age 55+

	(1) Income Decile 1	(2) Income Decile 2	(3) Income Decile 3	(4) Income Decile 4	(5) Income Decile 5	(6) Income Decile 6	(7) Income Decile 7	(8) Income Decile 8	(9) Income Decile 9	(10) Income Decile 10
Health Professional in Family	-0.008**	-0.017***	-0.016***	-0.013***	-0.016***	-0.007***	-0.009***	-0.009***	-0.008***	-0.005**
(no covariate)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
Health Professional in Family	-0.007**	-0.020***	-0.013***	-0.010***	-0.013***	-0.006**	-0.009***	-0.008***	-0.007***	-0.007***
(full covariates)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Mean of Dep. Var.	0.07	0.04	0.02	-0.00	-0.01	-0.02	-0.02	-0.03	-0.04	-0.06
Std. Dev. of Dep. Var.	0.66	0.63	0.60	0.57	0.57	0.55	0.54	0.53	0.51	0.48
Number of Observations	191,979	215,131	215,099	215,522	214,393	213,476	211,081	205,246	203,877	213,377

Panel C: HPV Vaccine, by age 20

	(1) Income Decile 1	(2) Income Decile 2	(3) Income Decile 3	(4) Income Decile 4	(5) Income Decile 5	(6) Income Decile 6	(7) Income Decile 7	(8) Income Decile 8	(9) Income Decile 9	(10) Income Decile 10
Health Professional in Family	0.073***	0.070***	0.062***	0.058***	0.059***	0.067***	0.063***	0.049***	0.033***	0.060***
(no covariate)	(0.009)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Health Professional in Family	0.047***	0.054***	0.048***	0.041***	0.053***	0.049***	0.048***	0.031***	0.017*	0.046***
(full covariates)	(0.009)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.008)
Mean of Dep. Var.	0.11	0.16	0.20	0.21	0.24	0.26	0.27	0.28	0.30	0.37
Std. Dev. of Dep. Var.	0.31	0.37	0.40	0.41	0.43	0.44	0.44	0.45	0.46	0.48
Number of Observations	11,327	12,568	12,516	12,469	12,539	12,427	12,511	12,500	12,471	13,683

Panel D: Tobacco Exposure, *in utero*

	(1) Income Decile 1	(2) Income Decile 2	(3) Income Decile 3	(4) Income Decile 4	(5) Income Decile 5	(6) Income Decile 6	(7) Income Decile 7	(8) Income Decile 8	(9) Income Decile 9	(10) Income Decile 10
Health Professional in Family	-0.066***	-0.076***	-0.069***	-0.066***	-0.058***	-0.056***	-0.051***	-0.041***	-0.028***	-0.015***
(no covariate)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Health Professional in Family	-0.046***	-0.033***	-0.025***	-0.022***	-0.012***	-0.012***	-0.007***	-0.003*	-0.000	0.001
(full covariates)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
Mean of Dep. Var.	0.33	0.28	0.24	0.21	0.19	0.18	0.17	0.15	0.12	0.08
Std. Dev. of Dep. Var.	0.47	0.45	0.43	0.40	0.39	0.38	0.38	0.36	0.32	0.27
Number of Observations	162,203	188,606	191,519	192,978	193,265	193,549	193,522	192,928	191,628	205,000

Notes: Tables report the results of OLS regressions of the outcome of interest on an indicator for having a health professional in the family, estimated separately for each decile (reported in columns (1) to (10)) of the individual's (or parental) income rank. In each panel, the first regression includes no covariates and the second regression includes the full set of covariates. Health professional in the family is an indicator variable that equals one if the individual has at least one relative with a completed medical or nursing degree. In panels A and B, the set of relatives includes spouse, sibling, cousin, child, child-in-law, niece/nephew, grandchild. In panels C and D, the set of relatives includes sibling, cousin, parent, aunt/uncle, and grandparent. Panel A restricts the sample to individuals born in Sweden in 1936 and 1937. Panel B restricts the sample to individuals born in Sweden between 1936-1961 and alive at age 55 and year 1997 (first year of inpatient claims). Panel C restricts the sample to females born in Sweden between 1995-1997 and alive at age 20. Panel D restricts the sample to children born in Sweden between 1995 and 2016. Covariates in Panels A and B include fixed effects for individual's own income rank percentile and income rank percentile of the highest-earning relative, year of birth fixed effects, gender dummy, fixed effects for discretized education levels, fixed effects for the county of residence at age 55. Covariates in panel C include fixed effects for parental income percentile at birth, highest-earning relative's income percentile, year of birth, gender, mother's county of residence in the year before child birth. Covariates in panel D include fixed effects for parental income percentile at birth, highest-earning relative's income percentile, year of birth, gender, mother's county of residence in the year before child birth, maternal birth order, mother's education, maternal age. Standard errors are clustered at the family level. *, **, and *** denote significance at 10%, 5%, and 1% level, respectively.

Table A2: Doctor in the Family and Health: Event Study Evidence

		Heterogeneity by					
		Income		Family Tie		Geographic Proximity	
	Pooled	Below Median	Above Median	Close	Far	Close	Far
Outcomes	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Heart Attack							
τ=-5 (τ: event year)	0.000	0.002	-0.001	0.001	0.001	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
τ=+10	-0.002*	-0.003*	-0.001	-0.002	-0.002	-0.002	-0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
τ=+15	-0.003***	-0.005**	-0.002	-0.004*	-0.003*	-0.003	-0.005***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Mean of Dep. Var. (at τ=+15)[a]	0.025	0.029	0.020	0.022	0.027	0.024	0.025
% Effect (at τ=+15)	12.0	17.2	10.0	18.2	11.1	12.5	20.0
Number of Observations	5,106,787	1,855,723	2,683,824	1,788,661	2,331,329	2,296,701	2,713,869
B. Heart Failure							
τ=-5	0.000	-0.000	0.000	-0.000	-0.000	-0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
τ=+10	-0.003***	-0.004**	-0.002**	-0.003*	-0.003**	-0.005***	-0.003**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
τ=+15	-0.005***	-0.005**	-0.003**	-0.005***	-0.005***	-0.006***	-0.005***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Mean of Dep. Var. (at τ=+15)[a]	0.022	0.025	0.015	0.019	0.025	0.022	0.021
% Effect (at τ=+15)	22.7	20.0	20.0	26.3	20.0	27.3	23.8
Number of Observations	5,106,787	1,855,723	2,683,824	1,788,661	2,331,329	2,296,701	2,713,869
C. Type II Diabetes							
τ=-5	0.003***	0.002	0.001	0.003*	0.003*	0.002*	0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
τ=+10	-0.007***	-0.005**	-0.006***	-0.007***	-0.005**	-0.008***	-0.006***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)
τ=+15	-0.010***	-0.006**	-0.010***	-0.011***	-0.008***	-0.013***	-0.008***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Mean of Dep. Var. (at τ=+15)[a]	0.044	0.049	0.034	0.039	0.048	0.046	0.043
% Effect (at τ=+15)	22.7	12.2	29.4	28.2	16.7	28.3	18.6
Number of Observations	5,106,787	1,855,723	2,683,824	1,788,661	2,331,329	2,296,701	2,713,869

Table A2: Doctor in the Family and Health: Event Study Evidence (Continued)

		Heterogeneity by					
		Income		Family Tie		Geographic Proximity	
	Pooled	Below Median	Above Median	Close	Far	Close	Far
Outcomes	(1)	(2)	(3)	(4)	(5)	(6)	(7)
D. Lung Cancer							
τ=-5	-0.000	0.000	-0.000	-0.000	0.000	-0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
τ=+10	-0.001***	-0.002***	-0.001	-0.002***	-0.001	-0.002***	-0.001**
	(0.000)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
τ=+15	-0.001*	-0.002	-0.001	-0.003**	-0.000	-0.002**	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Mean of Dep. Var. (at τ=+15)[a]	0.005	0.006	0.004	0.005	0.006	0.006	0.004
% Effect (at τ=+15)	20.0	33.3	25.0	60.0	0.0	33.3	0.0
Number of Observations	5,106,787	1,855,723	2,683,824	1,788,661	2,331,329	2,296,701	2,713,869

[a] Among family members of lawyers

Table reports coefficients σ_τ from the event study specification in Equation (4). The event time, sample restriction, and the set of family members included in the analysis are described in Section IID. Column (1) reports pooled results for 1936-1961 cohorts. Columns 2 and 3 split the sample by whether the individual's income rank within his/her gender-birth cohort is below or above the 50th percentile, with income measured at age 55. Individuals with a zero or negative income at age 55 are dropped from analyses. Columns 4 and 5 split the full sample by the type of family tie: parents-children in "close" family tie and aunts/uncles vs. nieces/nephews in "far." Individuals with both ties are excluded from analyses in Column 5. Columns 6 and 7 split the sample by geographic distance. Family members are classified as living "close" if their place of residence is recorded to be in the same county for more than 50 percent of the years between matriculation (into law or medicine) and the last year of data (2016). Coefficients are reported for event years -5, 10, and 15 (i.e. 5 years before, and 10, and 15 years after matriculation into the study of medicine or law). All regressions include the main effects and the interactions between event year dummies and the dummy for having a doctor in the (broad) family. The regressions further include the following covariates: age fixed effects, calendar year fixed effects, and individual fixed effects. Standard errors clustered by family are in parentheses. *, **, and *** denote significance at 10%, 5%, and 1% level, respectively.

Table A3: Survey Evidence on Health Literacy and Education

	Prefer Seeing Same Doctor (1)	Believe Doctor Always Tells Truth (2)	Regular Vegetables (3)	Regular Fruit (4)	Regular Sport (5)	Not Smoking (6)	Good Health (7)
No College Degree	0.06*	-0.07**	-0.19***	-0.16***	-0.06	-0.14***	-0.06**
	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)
Age Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Survey Weights Used	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Survey Year	2004	2004	2014	2014	2014	2014	2004
Mean of Dep. Var.	0.70	0.28	0.77	0.55	0.56	0.84	0.76
Std. Dev. of Dep. Var.	0.46	0.45	0.42	0.50	0.50	0.36	0.43
Number of Observations	927	927	738	738	738	738	927

Table reports OLS relationship between the level of education and health-related behaviors. The analysis is based on 2004 and 2014 waves of the European Social Survey for Sweden. The sample is restricted to working age individuals between age 30 and 60. We regress the outcome of interest on an indicator for having no college education (defined as not having a "completed tertiary education"). The OLS regression uses post-stratification survey weights and controls for age fixed effects. Binary outcome variables were constructed from underlying categorical survey responses to the following 7 questions or statements: 1) "Prefer same doctor for all everyday health problems"; 2) "Doctors keep whole truth from patients"; 3) "How often eat vegetables or salad, excluding potatoes"; 4) "How often eat fruit, excluding drinking juice"; 5) "Do sports or other physical activity, how many of last 7 days"; 6) "Cigarettes smoking behavior"; 7) "Subjective general health". For example, "Prefer seeing the same doctor" takes the value of one if individuals answered "Same for all health problems" in response to the question of whether they "Prefer same doctor for all everyday health problems." *, **, and *** denote significance at 10%, 5%, and 1% level, respectively.

A Identification codes for diseases and drug use

Diseases We identify diseases using the following ICD-10 diagnosis codes:

Conditions	ICD-10 Codes
Heart Attack	I21, I22, I23
Heart Failure	I11, I13, I50
Type II Diabetes	E11, E13, E14
Lung Cancer	C34
Addiction	F10-F19
Injury/Poisoning	S0-S9, T0-T9
Respiratory Infection	J00-J06, J20-J22
Intestinal Infection	A00-A09
Chronic Tonsil Diseases	J35
Asthma	J45
Hypertension	I10
Hyperlipidemia	E78, I70
Ischemic Heart Diseases	I20-I25
Stroke	I60, I61, I63, I66, G45, G46
Pregnancy, Childbirth and the Puerperium	O00-O99

Drug use Drug use is identified based on the following Anatomical Therapeutic Chemical (ATC) codes:

Drugs	ATC Codes
Statins	C10AA
Blood Thinners	B01AC
Diabetes Drugs	A10B
Beta Blockers	C07
Asthma Drugs	R03
Vitamin D	A11CB, A11CC
Hormonal Contraceptives	G03A excluding G03AD
HPV Vaccine	J07BM

B Study cohorts and data years

We use these health care records to construct variables that capture health outcomes and health investments at various points in the life cycle. We use different cohorts to study different outcomes. This is a natural consequence of four facts: First, different outcomes are observed (and relevant) at different points in the life cycle. For example, we study mortality by age 80; this requires studying individuals who are old enough for us to know whether they were alive at that age. Second, our different health records span slightly different years. We have in- and outpatient claims and birth records through 2016, but drug claims and death records through 2017. Third, the cohorts we use depend on the observation window in the outcome. For example, since the inpatient data range is 1995 through 2016, to observe the number of inpatient stays through age five, we need to restrict the sample to individuals born between 1995 and 2011. Fourth, the cohorts that we use are also driven by the years for which our key sources of variation are available.

	Study cohorts	Data years
Panel A. Descriptive analysis		
Mortality by age 80	Birth cohorts 1936-1937	1961-2017
Lifestyle-related conditions	Birth cohorts 1936-1961	1997-2016
HPV vaccination	Birth cohorts 1995-1997, females	2005-2017
Tobacco exposure	Birth cohorts 1995-2016	1995-2016
Asthma, by age 5	Birth cohorts 2001-2011	2001-2016
High-risk birth	Birth cohorts 1995-2016	1995-2016
Number of inpatient stays, by age 5	Birth cohorts 1995-2011	1995-2016
Number of inpatient stays, age 45-50	Birth cohorts 1950-1966	1995-2016
Panel B. Lottery analysis		
Older family members		
Diseases/Preventable hospitalizations	Family members aged 50 or above of application cohorts 2007 fall-2008 fall[a]	2008-2016
Drug use other than vitamin D	Family members aged 50 or above of application cohorts 2007 fall-2009 fall[a]	2008-2017
Vitamin D	Female family members aged 50 or above of application cohorts 2007 fall-2009 fall[a]	2008-2017
Younger family members		
Outcomes except HPV vaccine/contraceptives	Family members aged below 25 and born before the application of application cohorts 2007 fall-2010 spring[a]	2008-2016
HPV vaccination	Female family members aged 10-25 of application cohorts 2007 fall-2010 spring[b]	2008-2017
Hormonal contraceptives	Female family members aged 10-20 of application cohorts 2007 fall-2010 spring[b]	2008-2017
Panel C. Event study		
Mortality	Birth cohorts 1936-1940	1961-2017
Lifestyle-related conditions	Birth cohorts 1936-1961	1997-2016

[a] Age refers to age at the year of the applicant's medical school application.

[b] Age refers to age at the end of the tracking period, i.e., six years after the applicant's medical school application.

C Examining zero-sum, and thus conceptually non-scalable, benefits of exposure to experts

To get a sense of how the "social capital" channel could operate and how likely it is to be quantitatively significant in our context, we discuss several examples of this channel on outcomes that we can capture in the data that reflect expensive or (in our setting) potentially rationed health care services.

First, we investigate whether family members of health professionals obtain more expensive heart attack treatments. The underlying idea is as follows. There are two common invasive therapies, one of which is substantially more expensive than the other, and one non-invasive (drug) therapy, which is the cheapest; under the "social capital" hypothesis, connected patients may be more likely to get a relatively expensive treatment option (holding severity of the condition constant).[1] We find no evidence of differences in the probability of getting an invasive (versus non-invasive) heart attack treatment across patients with and without a health professional in the extended family. Further, we find no difference in the intensity of invasive treatment conditional on getting an invasive (i.e. surgery rather than drugs) treatment.

Second, we investigate whether family members of health professionals have systematically longer stays in the hospital after childbirth (conditioning on a wide range of characteristics capturing postpartum maternal health and the child's health at birth). Despite the fact that the duration of postpartum care is generally rationed in the Swedish health care system – mothers are discharged as early as six hours after childbirth, while mothers in the U.S. are legally entitled to stay for up to 48 hours, depending on the state – we do not find any differences in the length of stay across patients with and without a health professional in the family.

Third, we examine the importance of the access to care channel for cancer treatment, as existing literature has documented that connections appear relevant for the choice and speed of cancer treatments Fiva et al. (2014). Here, we do find a smaller time window between the first diagnosis of breast cancer and breast cancer surgery among family members of health professionals. Interestingly, however, there is no pronounced income gradient in the prevalence of cancers, nor in mortality attributable to cancer.[2] This suggests that, at the population level, the access to care channel does not generate substantial differences in cancer-related outcomes across the income distribution.

In sum, the broader picture that emerges is that, while some non-scalable benefits may be at play in our empirical setting, our results allow us to argue that there *exists* an effect of exposure to expertise that does not run through social capital, that may be scalable, and that is currently being left on the table.[3]

[1] The two invasive therapies are coronary artery bypass grafting (CABG) and percutaneous coronary intervention (PCI), with the former being a more expensive open heart surgery.

[2] In fact, if anything, we observe an *inverse* SES-gradient in cancers, which likely is driven by competing risks with cardiovascular diseases as well as more screening at the upper end of the income distribution. The exception is lung cancer; however, it accounts for a very small share of all cancers.

[3] Our results are consistent with the idea that historically, the "extensive" margin of disease prevention through changes in social and individual investments into health - that we hypothesize can be affected by access to health expertise - has had a substantially more pronounced effect on population health than the "intensive" margin of moving from a lower to a higher quality of care provider, or getting care faster within a given system (Cutler and Lleras-Muney 2008).

D Further related literature

D.1 Early childhood interventions

A growing literature has documented that early life interventions have a positive effect on infant mortality and can promote health in the long-run, suggesting that conditions in infancy are a relevant source of health and socioeconomic disparities in later life. For a universal home visiting program implemented in Denmark, Wüst (2012) show that the intervention had a positive and significant effect on the infant first-year survival rate in Danish towns and was most effective in the majority of small and medium-sized municipalities. The authors suggests that the main driver of the program's impact was the promotion of breastfeeding and appropriate infant nutrition by visiting nurses. In a related study, Hjort et al. (2017) examine the long-term impact of the Danish home visiting program. They find that treated individuals that were visited by nurses in infancy experience better health mid-life: they have lower mortality rates, spend fewer nights at hospital, and are less likely to be diagnosed with cardiovascular diseases. Similarly, Bütikofer et al. (2019) investigate the long-term impact of mother and child health care centers in Norway and find that the increasing access to well-child visits had a positive effect on health, education and earning of treated infants when they reach age 30 to 40. Moreover, the authors find a stronger impact for children from lower socioeconomic background. Similarly, Sweden saw the introduction of a nurse home visiting program in the early 1930's and Bhalotra et al. (2017) find that, in the long-run, the infant care provided by nurse home visits reduced the probability of dying by age 75 by seven percent.

D.2 Community health workers and access to primary care

Community health workers (CHWs) have been employed in many countries to provide health-related services to their fellow community members. Although there has not been many rigorous evaluations, most existing evidence suggests that CHWs increase takeup rates of a wide variety of healthy behaviors and improve disease management in the community, notably for health behaviors such as cancer screening and immunization, and management of diseases such as asthma, hypertension, and diabetes (see e.g., Norris et al. 2006, Haines et al. 2007, and Najafizada et al. 2015). In addition, by assisting individuals in navigating the health care system, CHWs have been shown to improve access to medical services, especially for marginalized populations (Felix et al. 2011, Najafizada et al. 2015).

Access to primary care in the community setting has also been found to be an effective way to improve patients' health. Bailey and Goodman-Bacon (2015) use the rollout of community health centers (CHCs) in the U.S. from 1965 to 1974 to study the long-term health benefits of increased access to primary health care for the poor. The paper finds that, in one decade after CHCs were established, CHCs reduced age-adjusted all-cause mortality rates by 7 to 13 percent among the poor aged 50 and older, with the reduction primarily driven by the decline in cardiovascular-related deaths. The authors argue that having access to a regular source of care, lower medication cost, and improved compliance with prescription drugs were the main mechanisms for the effects of

CHCs on mortality.

Moreover, a growing body of evidence suggests that the ease of access to nurses improves the health of patients with chronic conditions. Fergenbaum et al. (2015) present a systematic review of six randomized control trials that study the effects of home visits with nurse-led guidance in disease self-care management. Home visiting programs result in fewer hospitalizations, fewer emergency department visits, and better patient quality of life. Studies on nurse-led clinics that provide disease knowledge and support for disease self-care management report similar health effects: these clinics significantly reduce patient emergency department visits, hospital readmissions, and mortality rates, and improve patient medication adherence (Agvall et al. 2013, Gandhi et al. 2017, and Liljeroos and Strömberg 2019).

D.3 Patient education

An extensive body of work has evaluated the effectiveness of patient education interventions, finding such programs to be generally effective in promoting population health. For chronic diseases, Stenberg et al. (2018) review existing studies - 56 face-to-face intervention among patients living with chronic illess - on the impact of education programs that target chronic obstructive pulmonary disease (COPD), asthma, chronic pain, heart disease, and diabetes patients. The authors find that, regardless of study design and time horizon, interventions that promote patient education are beneficial in terms of decreased hospital admissions, fewer visits to emergency departments or general practitioners, and in terms of increased quality-adjusted life-years

Similarly, Wang et al. (2017) review randomized control trials that investigate effects of self-management education among patients with COPD. The paper highlights that such education programs improve patient disease-specific knowledge and quality of life, and reduce respiratory-related hospital admissions and emergency department visits. Anderson et al. (2017) focus on the educational component of cardiac rehabilitation for patients with coronary heart diseases. The study reviews 22 randomized control trials that assigned patients to different educational interventions that ranged from face-to-face counseling to residential stays with follow-up sessions. Patients in control groups received usual medical care in cardiac rehab that comprises exercise counseling and training and psychological support. The paper finds that, although there is limited evidence that education-based interventions reduce total mortality, the risk of a heart attack, or the number of hospitalizations, these interventions result in lower risks of cardiovascular events and better quality of life. Similarly, Menichetti et al. (2018) review randomized control trials that promote patient engagement among older adults with osteoporosis, diabetes or cardiovascular-related health problems. The authors find that such interventions often demonstrate positive effects on patient compliance with treatment regimens.

In the context of health behaviors, Aveyard et al. (2012) and Stead et al. (2013) show that medical advice and provision of behavioral or pharmaceutical assistance on smoking cessation increase the frequency and success of smoking cessation attempts. Kaner et al. (2018) review the literature on alcohol interventions provided by health professionals and conclude positive effects of these interventions on reducing excessive alcohol consumption. For weight control, Aveyard et al. (2016) show that a randomized trial that provides interventions delivered by trained general practitioners improves body weight control among obese patients.

Another strand of literature examines the effects of public health education campaigns promoted by social media. A comprehensive summary of this literature can be found in Giustini et al. (2018). Many topics have been included in these social-media education campaigns, including health behaviors such as smoking cessation, healthy diet and physical activity (see, e.g., Chang et al. 2013, Williams et al. 2014, Swanton et al. 2015, and Chakraborty et al. 2018), and prevention and management of diseases such as diabetes and cancer (Gabarron et al. 2018, Han et al. 2018). Existing studies generally suggest positive effects of these campaigns on population health.

E Interpreting magnitudes in "doctor-equivalents"

We consider a stylized setting that divides individuals in our sample into two groups – those in the top and bottom halves of the income distribution, respectively. In the data, the average probability of dying by age 80, conditional on having survived until age 55, is 35 percent in the first half of the income distribution and 27 percent in the second half of the income distribution. These two data moments are plotted as a solid line in Panel A of Appendix Figure A10. From our estimates of treatment effects in Appendix Table A1 and Table 4, we have that exposure to a physician in the family leads to a 10 percent decline in mortality (on average). If the health benefits from exposure to expertise scaled linearly with each doctor in the family, it would take a difference of 2.1 "doctor-equivalents" on average between families at the top and the bottom of the income distribution to account for the full difference in mortality.[4]

Naturally, the *levels* of intra-family exposure that we observe in the data are much lower – in practice, exposure to doctors happens not only through family members, but also through friends, neighbors, and colleagues. What's more, health literacy does not only stem from relatives who are health professionals, but from a range of sources, all of which are likely to be more readily available at the top of the income distribution (Kindig et al. 2004). The gradient in exposure between the first and second half of the income distribution that we observe in the data, however, is striking: While individuals at the top half of the income distribution have on average 0.26 physicians in the family, those at the bottom have on average 0.05, or five times fewer, doctors in the family. To get 2 more "doctor-equivalents" at the top of the income distribution, keeping the gradient the same, we would need to have at least 2.5 "doctor-equivalents" at the top of the income distribution on average and 0.5 at the bottom. Another way to summarize these magnitudes, staying for simplicity with exposure to physicians per se, is as follows. If individuals in the top half of the income distribution on average know one doctor, while only every fifth individual knows a doctor at the bottom of the income distribution, then this difference alone could account for a third of the health-income gradient.

[4] The "doctor-equivalents" required to fully close the gap would naturally be higher if the effect of an extra physician on health literacy and the effect of extra heath literacy had decreasing marginal returns, as is likely to be the case in practice.

Appendix references

Agvall, Björn, Urban Alehagen, and Ulf Dahlström, "The Benefits of Using a Heart Failure Management Programme in Swedish Primary Healthcare," *European Journal of Heart Failure*, 2013, *15* (2), 228–236.

Anderson, Lindsey, James Pr Brown, Alexander M Clark, Hasnain Dalal, Henriette K Rossau, Charlene Bridges, and Rod S Taylor, "Patient Education in the Management of Coronary Heart Disease," *Cochrane Database of Systematic Reviews*, June 2017, *6*, No. CD008895.

Aveyard, Paul, Amanda Lewis, Sarah Tearne, Kathryn Hood, Anna Christian-Brown, Peymane Adab, Rachna Begh, Kate Jolly, Amanda Daley, Amanda Farley, Deborah Lycett, Alecia Nickless, Ly-Mee Yu, Lise Retat, Laura Webber, Laura Pimpin, and Susan Jebb, "Screening and Brief Intervention for Obesity in Primary Care: A Parallel, Two-Arm, Randomised Trial," *The Lancet*, 2016, *388* (10059), 2492–2500.

_, Rachna Begh, Amanda Parsons, and Robert West, "Brief Opportunistic Smoking Cessation Interventions: A Systematic Review and Meta-Analysis to Compare Advice to Quit and Offer of Assistance," *Addiction*, 2012, *107* (6), 1066–1073.

Bailey, Martha J and Andrew Goodman-Bacon, "The War on Poverty's Experiment in Public Medicine: Community Health Centers and the Mortality of Older Americans," *American Economic Review*, 2015, *105* (3), 1067–1104.

Bhalotra, Sonia, Martin Karlsson, and Therese Nilsson, "Infant Health and Longevity: Evidence from a Historical Intervention in Sweden," *Journal of the European Economic Association*, 2017, *15* (5), 1101–1157.

Bütikofer, Aline, Katrine V. Løken, and Kjell G. Salvanes, "Infant Health Care and Long-Term Outcomes," *The Review of Economics and Statistics*, 05 2019, *101* (2), 341–354.

Chakraborty, Bibhas, Raju Maiti, and Victor J Strecher, "The Effectiveness of Web-Based Tailored Smoking Cessation Interventions on the Quitting Process (Project Quit): Secondary Analysis of a Randomized Controlled Trial," *Journal of Medical Internet Research*, 2018, *20* (6), e213.

Chang, Tammy, Vineet Chopra, Catherine Zhang, and Susan J Woolford, "The Role of Social Media in Online Weight Management: Systematic Review," *Journal of Medical Internet Research*, 2013, *15* (11), e262.

Cutler, David and Adriana Lleras-Muney, "Education and Health: Evaluating Theories and Evidence," in Robert F. Schoeni, James S. House, George A. Kaplan, and Harold Pollack, eds., *Making Americans Healthier: Social and Economic Policy as Health Policy*, New York: Russell Sage Foundation, 2008.

European Social Survey, "European Social Survey Round 2 Data," 2004. Data file edition 3.6. NSD – Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC. doi:10.21338/NSD-ESS2-2004.

_ , "European Social Survey Round 7 Data," 2014. Data file edition 2.2. NSD – Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC. doi:10.21338/NSD-ESS7-2014.

Felix, Holly C, Glen P Mays, M Kathryn Stewart, Naomi Cottoms, and Mary Olson, "Medicaid Savings Resulted When Community Health Workers Matched Those with Needs to Home and Community Care," *Health Affairs*, 2011, *30* (7), 1366–1374.

Fergenbaum, Jennifer, Sarah Bermingham, Murray Krahn, David Alter, and Catherine Demers, "Care in the Home for the Management of Chronic Heart Failure: Systematic Review and Cost-Effectiveness Analysis," *Journal of Cardiovascular Nursing*, 2015, *30* (4S), S44–S51.

Fiva, Jon H, Torbjørn Hægeland, Marte Rønning, and Astri Syse, "Access to Treatment and Educational Inequalities in Cancer Survival," *Journal of Health Economics*, 2014, *36*, 98–111.

Gabarron, Elia, Eirik Årsand, and Rolf Wynn, "Social Media Use in Interventions for Diabetes: Rapid Evidence-Based Review," *Journal of Medical Internet Research*, 2018, *20* (8), e10303.

Gandhi, Sumeet, Wassim Mosleh, Umesh C Sharma, Catherine Demers, Michael E Farkouh, and Jon-David Schwalm, "Multidisciplinary Heart Failure Clinics are Associated with Lower Heart Failure Hospitalization and Mortality: Systematic Review and Meta-Analysis," *Canadian Journal of Cardiology*, 2017, *33* (10), 1237–1244.

Giustini, Dean, Syed Mustafa Ali, Matthew Fraser, and Maged N Kamel Boulos, "Effective Uses of Social Media in Public Health and Medicine: A Systematic Review of Systematic Reviews," *Online Journal of Public Health Informatics*, 2018, *10* (2), e215.

Haines, Andy, David Sanders, Uta Lehmann, Alexander K Rowe, Joy E Lawn, Steve Jan, Damian G Walker, and Zulfiqar Bhutta, "Achieving Child Survival Goals: Potential Contribution of Community Health Workers," *The Lancet*, 2007, *369* (9579), 2121–2131.

Han, Claire Jungyoun, Young Ji Lee, and George Demiris, "Interventions Using Social Media for Cancer Prevention and Management: A Systematic Review," *Cancer Nursing*, 2018, *41* (6), E19–E31.

Hjort, Jonas, Mikkel Sølvsten, and Miriam Wüst, "Universal Investment in Infants and Long-Run Health: Evidence from Denmark's 1937 Home Visiting Program," *American Economic Journal: Applied Economics*, 2017, *9* (4), 78–104.

Kaner, Eileen FS, Fiona R Beyer, Colin Muirhead, Fiona Campbell, Elizabeth D Pienaar, Nicolas Bertholet, Jean B Daeppen, John B Saunders, and Bernard Burnand, "Effectiveness of Brief

Alcohol Interventions in Primary Care Populations," *Cochrane Database of Systematic Reviews*, 2018, *2*, No. CD004148.

Kindig, David A, Allison M Panzer, and Lynn Nielsen-Bohlman, *Health literacy: A Prescription to End Confusion*, National Academies Press, 2004.

Liljeroos, Maria and Anna Strömberg, "Introducing Nurse-Led Heart Failure Clinics in Swedish Primary Care Settings," *European Journal of Heart Failure*, 2019, *21* (1), 103–109.

Menichetti, Julia, Guendalina Graffigna, and Aslak Steinsbekk, "What are the Contents of Patient Engagement Interventions for Older Adults? A Systematic Review of Randomized Controlled Trials," *Patient Education and Counseling*, 2018, *101* (6), 995–1005.

Najafizada, Said Ahmad Maisam, Ivy Lynn Bourgeault, Ronald Labonte, Corinne Packer, and Sara Torres, "Community Health Workers in Canada and other High-Income Countries: A Scoping Review and Research Gaps," *Canadian Journal of Public Health*, 2015, *106* (3), 157–164.

Norris, Susan L, Farah M Chowdhury, K Van Le, Tanya Horsley, J Nell Brownstein, Xuanping Zhang, Leonard Jack Jr, and Dawn W Satterfield, "Effectiveness of Community Health Workers in the Care of Persons with Diabetes," *Diabetic Medicine*, 2006, *23* (5), 544–556.

Stead, Lindsay F, Diana Buitrago, Nataly Preciado, Guillermo Sanchez, Jamie Hartmann-Boyce, and Tim Lancaster, "Physician Advice for Smoking Cessation," *Cochrane Database of Systematic Reviews*, 2013, *5*, No. CD000165.

Stenberg, Una, Andre Vågan, Maria Flink, Vibeke Lynggaard, Kari Fredriksen, Karl Fredrik Westermann, and Frode Gallefoss, "Health Economic Evaluations of Patient Education Interventions: A Scoping Review of the Literature," *Patient Education and Counseling*, 2018, *101* (6), 1006–1035.

Swanton, Rosie, Vanessa Allom, and Barbara Mullan, "A Meta-Analysis of the Effect of New-Media Interventions on Sexual-Health Behaviours," *Sexually Transmitted Infections*, 2015, *91* (1), 14–20.

The Health Inequality Project, "Data Tables." Accessed January 2019. `https://healthinequality.org/`.

Wang, Tao, Jing-Yu Tan, Lily Dongxia Xiao, and Renli Deng, "Effectiveness of Disease-Specific Self-Management Education on Health Outcomes in Patients with Chronic Obstructive Pulmonary Disease: An Updated Systematic Review and Meta-Analysis," *Patient Education and Counseling*, 2017, *100* (8), 1432–1446.

Williams, Gillian, Michele P Hamm, Jocelyn Shulhan, Ben Vandermeer, and Lisa Hartling, "Social Media Interventions for Diet and Exercise Behaviours: A Systematic Review and Meta-Analysis of Randomised Controlled Trials," *BMJ Open*, 2014, *4* (2), e003926.

Wüst, Miriam, "Early Interventions and Infant Health: Evidence from the Danish Home Visiting Program," *Labour Economics*, 2012, *19* (4), 484–495.