Decomposition Techniques for Social Epidemiology

S Harper

PhD course: Advanced Social Epidemiology, 19th Aug – 23rd Aug, 2019

University of Copenhagen

Overview of Decomposition Techniques

Today:

- Life table decomposition
- Inequality decomposition: Concentration Index
- Decomposing two-group differences: Blinder-Oaxaca

Not covered today, but of interest:

- Effect decomposition (i.e., mediation)
- Decomposition of population rates
- Inequality decomposition: Indexes for Nominal social groups

Decomposition: Moving from Description to Explanation

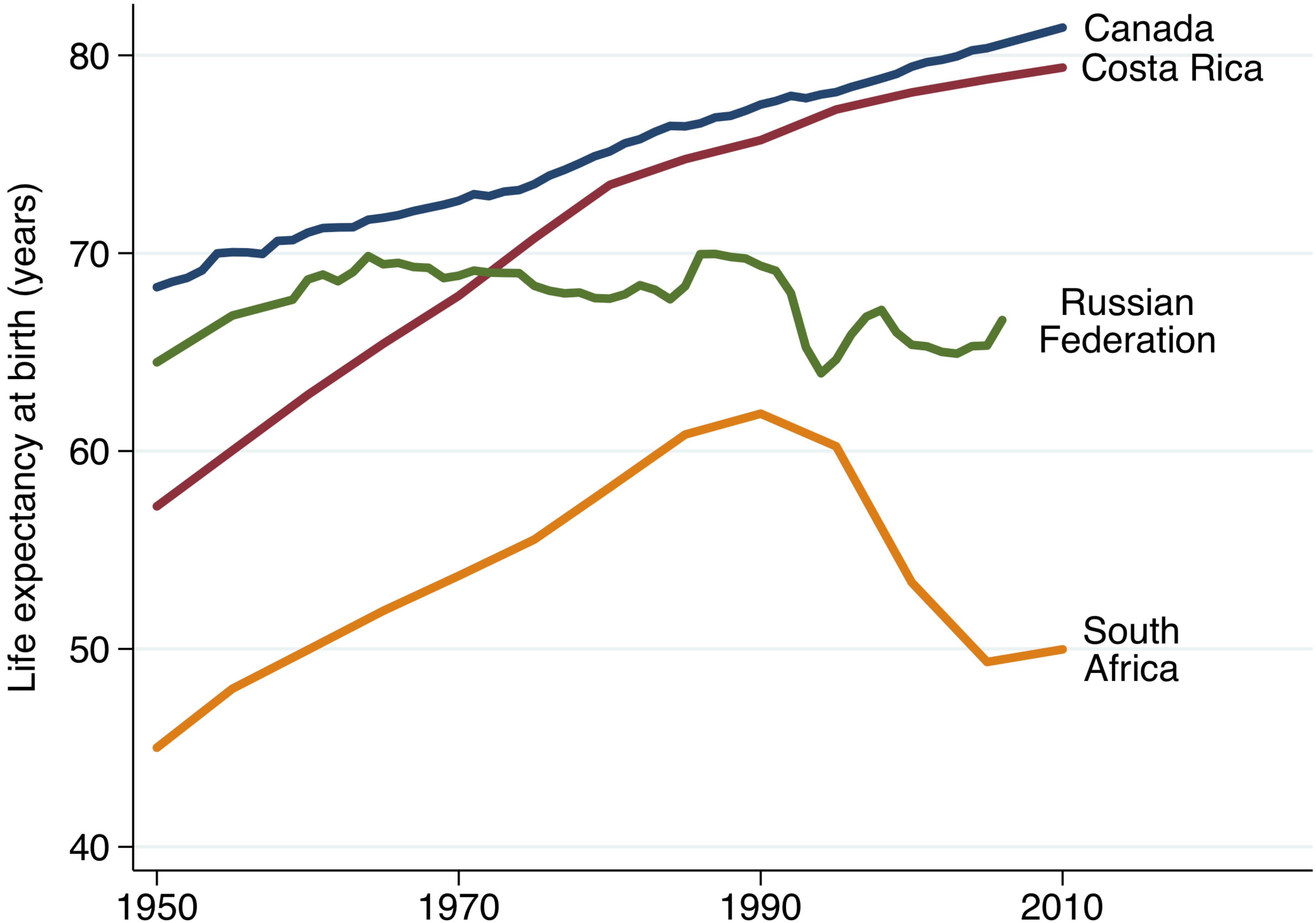
- Ultimately, we want to know why health inequalities are changing over time—what changed?
  - Risk factors?
  - Demographic composition?
  - Social conditions?

- Unpacking the ‘components’ of health inequality is an opportunity to better integrate the monitoring of health inequalities with the etiology of health inequalities.

- These techniques often involve various kinds of ‘counterfactual’ scenarios

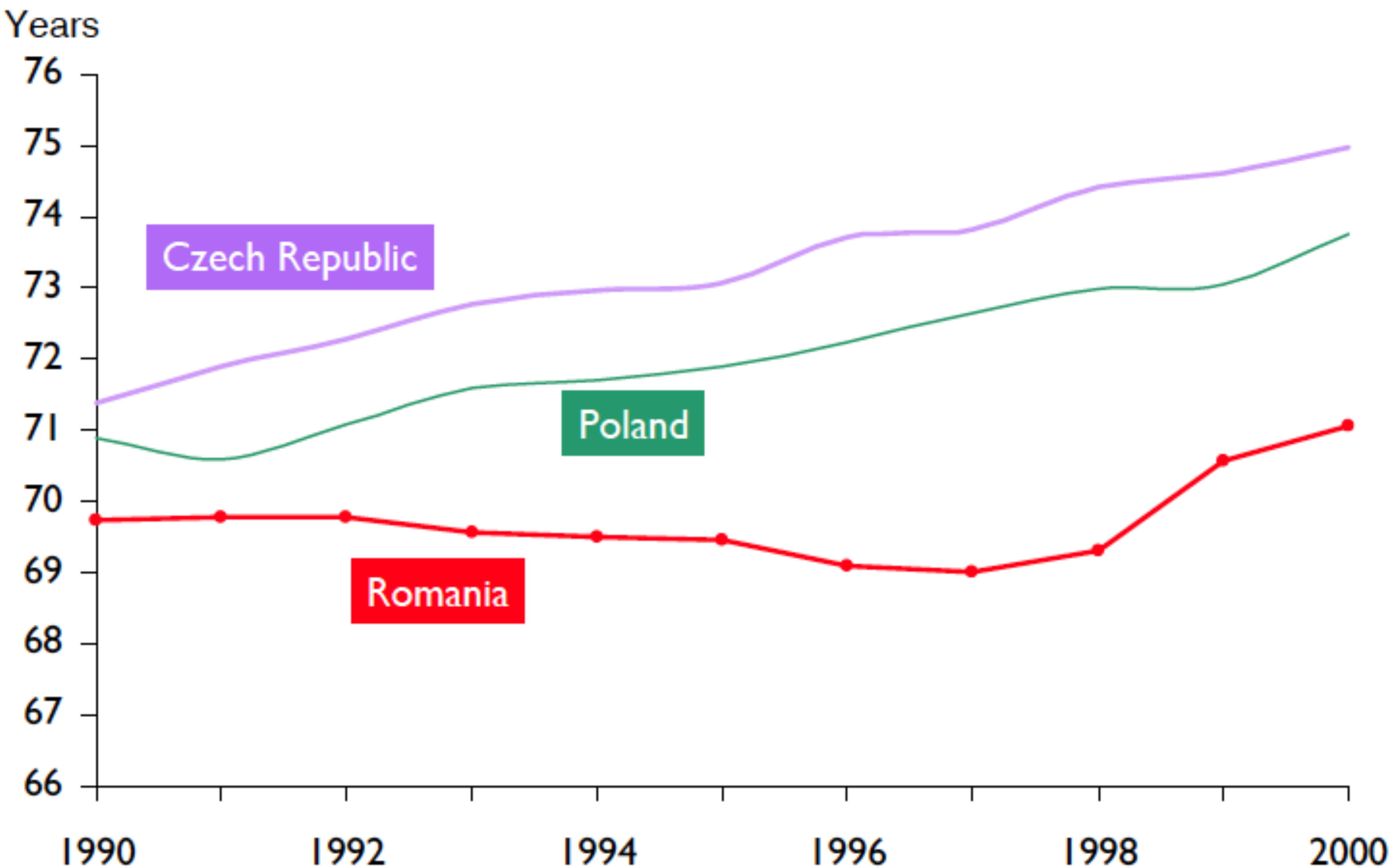
Life Table Decomposition

Why Does Life Expectancy Go Up (and Down)?

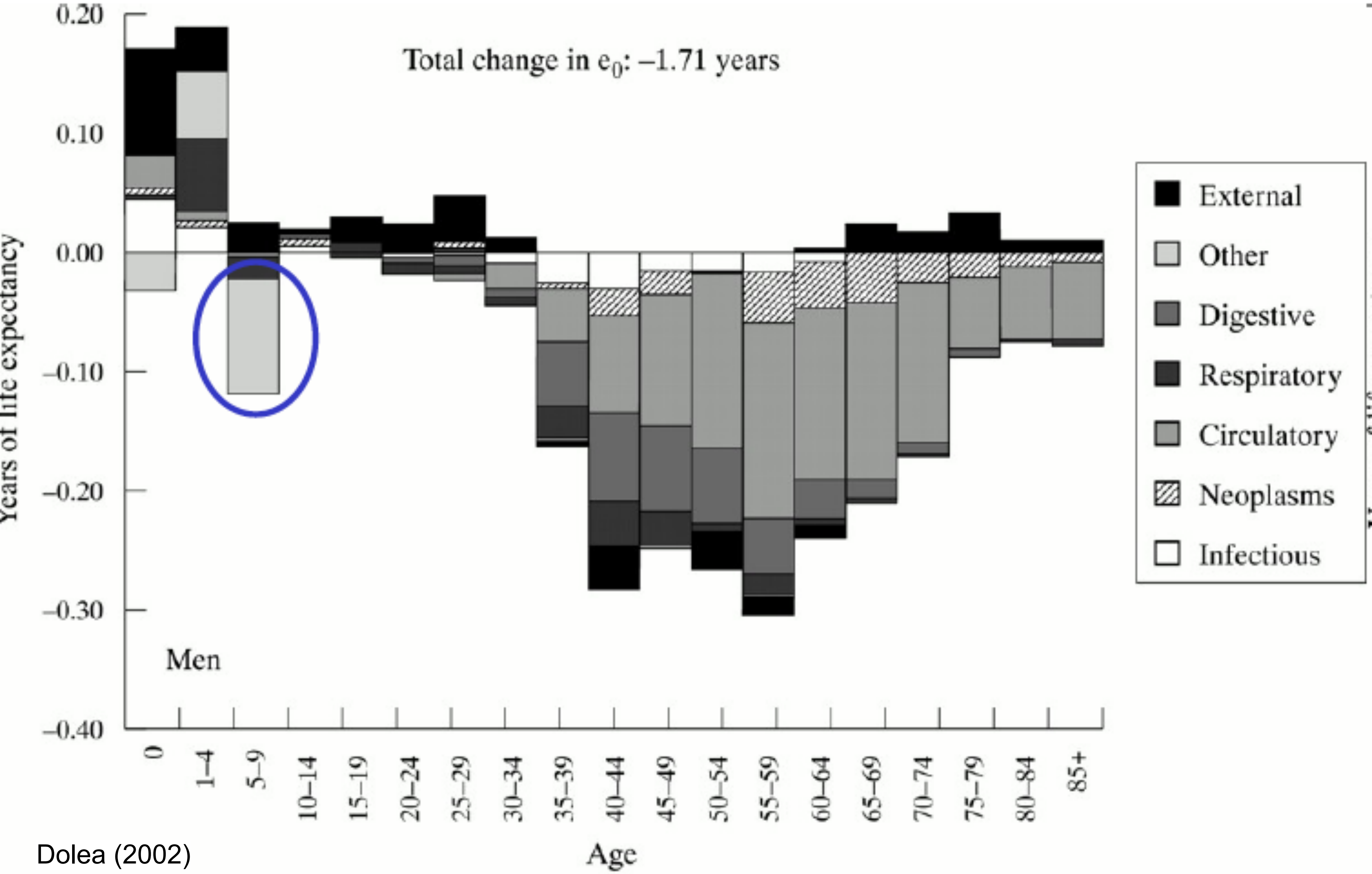


Source: GapMinder, 2010

Life Expectancy in Former Communist Countries

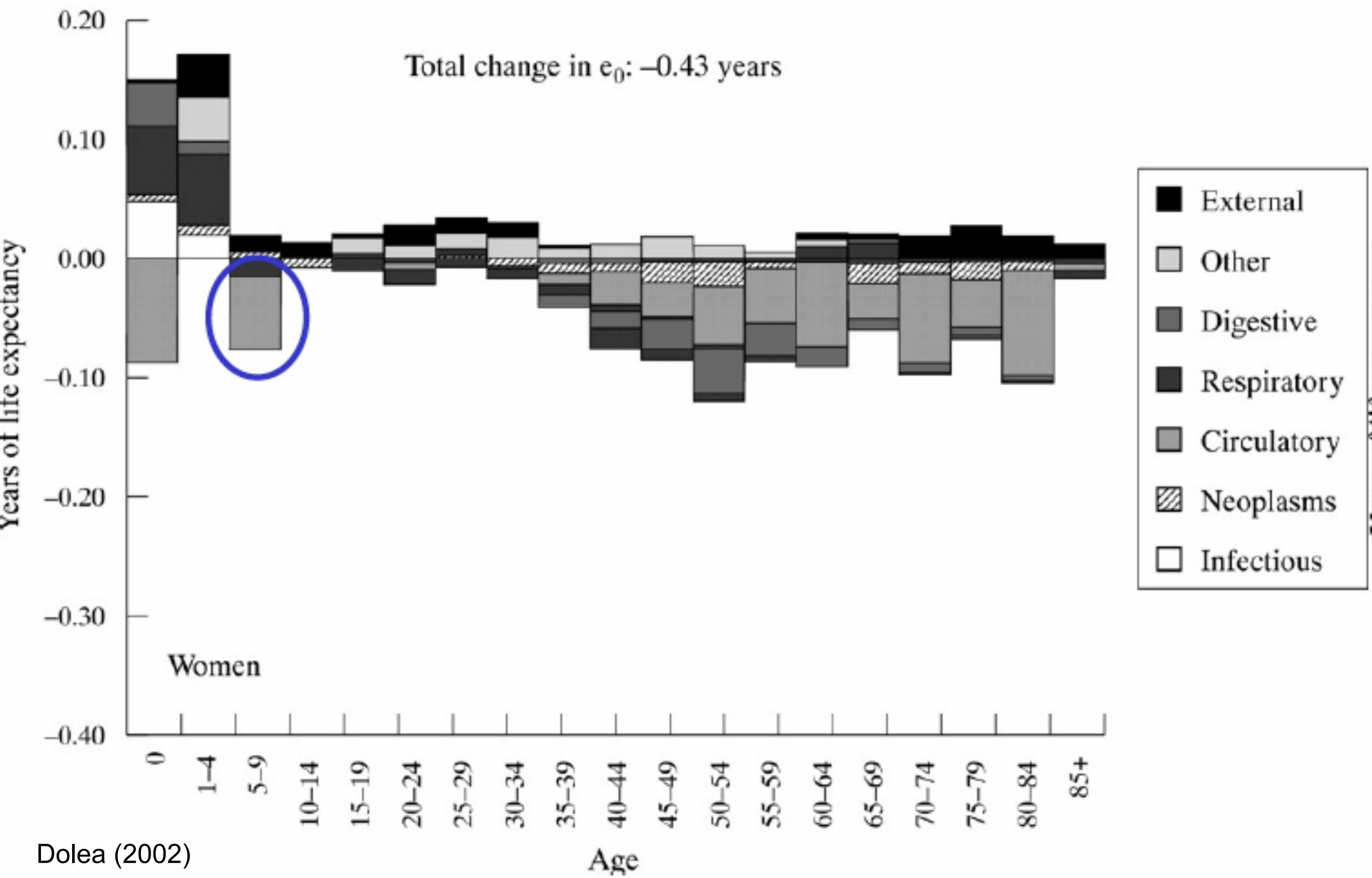


Contribution of Causes of Death to Romanian LE Change - Men



Dolea (2002)

Contribution of Causes of Death to Romanian LE Change - Women



Dolea (2002)

Linking Decomposition to Specific Health Determinants

“One factor, HIV infection, has behaved in a very specific way in Romania. As we have shown, the increase in deaths among those aged 5–9 is almost entirely attributable to AIDS. In the late 1980s, over 100,000 children were living in so called "orphanages" in Romania. "Micro-transfusions" of blood were commonly administered as treatment for anaemia or malnutrition, in many cases leading to HIV infection.”

Dolea (2002)

Published by Oxford University Press on behalf of the International Epidemiological Association

International Journal of Epidemiology 2010;1–9
doi:10.1093/ije/dyq073

Trends in the gap in life expectancy between Arabs and Jews in Israel between 1975 and 2004

Wasef Na'amnih,[1] Khitam Muhsen,[1,2] Jalal Tarabeia,[3] Ameed Saabneh[4] and Manfred S Green[1,3]*

Understanding the Rapid Increase in Life Expectancy in South Korea

Seungmi Yang, PhD, Young-Ho Khang, MD, PhD, Sam Harper, PhD, George Davey Smith, MD, DSc, David A. Leon, PhD, and John Lynch, PhD

ORIGINAL ARTICLE

CAUSES AND CONTRIBUTIONS TO DIFFERENCES IN LIFE EXPECTANCY FOR INUIT NUNANGAT AND CANADA, 1994–2003

Paul A. Peters

Health Analysis Division, Statisti

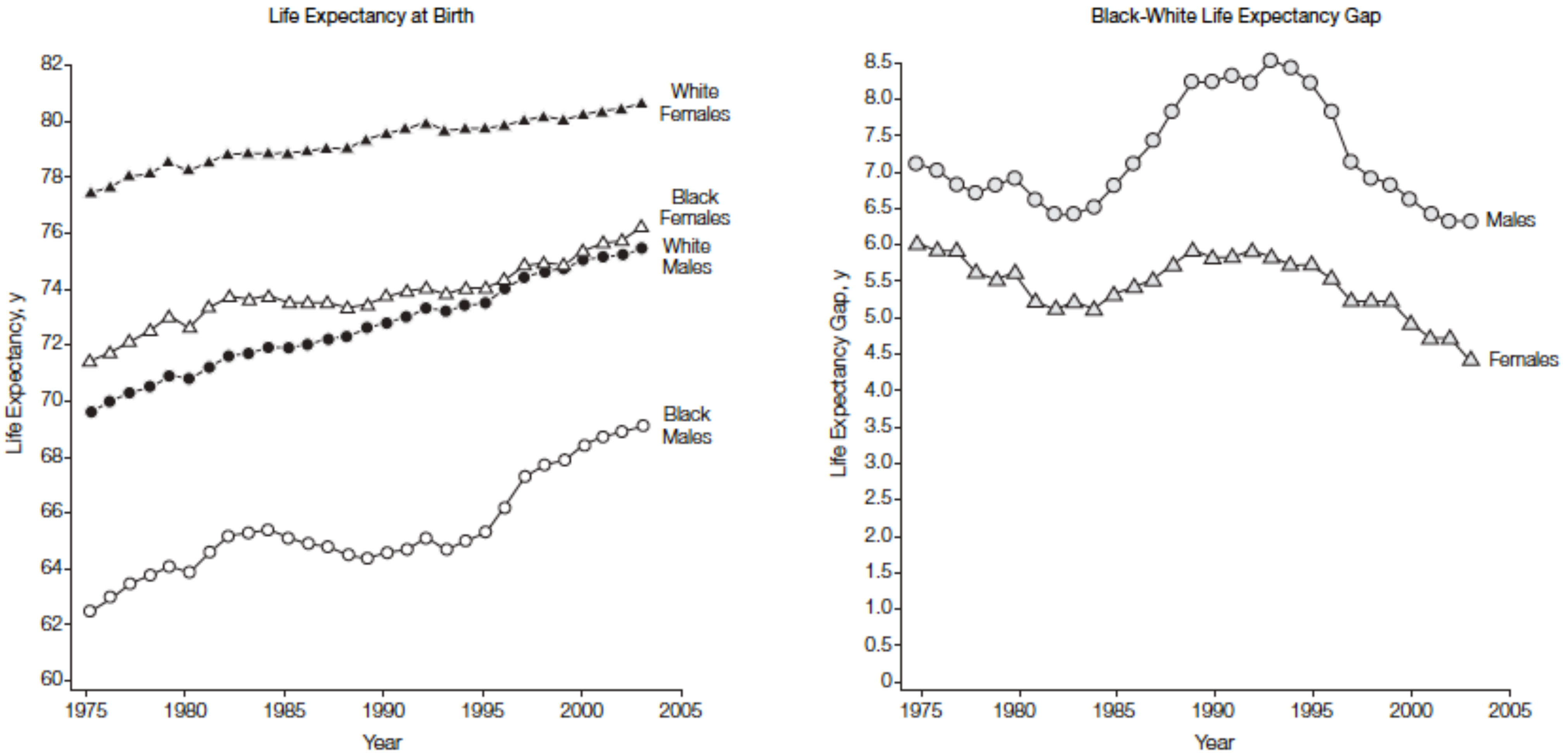
Received 10 August 2009; Accep

Letter From Russia

Causes of Declining Life Expectancy in Russia

Francis C. Notzon, PhD; Yuri M. Komarov, MD; Sergei P. Ermakov, PhD; Christopher T. Sempos, PhD; James S. Marks, MD; Elena V. Sempos, MD

Figure 1. Life Expectancy at Birth Among Black and White Males and Females in the United States and the Black-White Life Expectancy Gap, 1975-2003



Data taken from the United States Life Tables of the National Center for Health Statistics.[7]

Life Expectancy at Birth (years), United States

	Women			Men		
Year	White	Black	Gap	White	Black	Gap
1993	79.78	74.19	5.59	73.20	64.77	8.44
2003	80.49	75.94	4.54	75.42	69.09	6.33
Δ93-03	0.71	1.75	-1.04	2.22	4.32	-2.11

US Life Table for Black Females, 2003

Age	Length of interval	Probability of dying between ages x to x+n	Number surviving to age x	Number dying between ages x to x+n	Person-years lived between ages x to x+n	Total number of person-years lived above age x	Life exp at age x
x	n	$_nq_x$	$_nl_x$	$_nd_x$	$_nL_x$	T_x	e_x
0	1	0.0123	100,000	1,229	98,900	7,594,342	75.94
1	4	0.0016	98,771	155	394,698	7,495,442	75.89
5	5	0.0009	98,616	88	492,842	7,100,744	72.00
10	5	0.0010	98,528	98	492,389	6,607,902	67.07
15	5	0.0019	98,430	187	491,758	6,115,513	62.13
20	5	0.0035	98,243	345	490,362	5,623,755	57.24
25	5	0.0047	97,898	460	488,415	5,133,394	52.44
35	10	0.0105	96,794	1,021	481,552	4,159,267	42.97
45	10	0.0242	94,229	2,277	465,727	3,202,492	33.99
55	10	0.0483	88,782	4,287	433,781	2,284,543	25.73
65	10	0.0976	78,537	7,662	374,209	1,442,517	18.37
75	10	0.2024	60,885	12,321	274,487	738,005	12.12
85	∞	1.0000	34,617	34,617	255,202	255,202	7.37

"Decomposing" Life Expectancy Difference, by Age

	Black Females, 1993				White Females, 1993			
x	l_x	${}_nL_x$	T_x	e_x	l_x	${}_nL_x$	T_x	e_x
0-1	100,000	98,702	7,418,554	74.2	100,000	99,464	7,977,549	79.8
1-4	98,552	393,515	7,319,853	74.3	99,398	397,269	7,878,085	79.3
5-9	98,279	490,998	6,926,338	70.5	99,265	496,109	7,480,816	75.4
:								
85+	30,497	232,215	232,214	7.6	42,768	296,729	296,729	6.9

Direct effect | Indirect effect + interaction

$$ {}_n\Delta_x = \left[\frac{l_x^{Black}}{l_0^{Black}} \times \left(\frac{{}_nL_x^{White}}{l_x^{White}} - \frac{{}_nL_x^{Black}}{l_x^{Black}} \right) \right] + \left[\frac{T_{x+n}^{White}}{l_{x+n}^{White}} \times \frac{\dfrac{l_x^{Black}\, l_{x+n}^{White}}{l_x^{White}} - l_{x+n}^{Black}}{l_0^{Black}} \right] $$

Fraction of survivors at each age

Difference in person-years lived within age group "temporary life expectancy"

(Arriaga, 1984)

"Decomposing" Life Expectancy Difference, by Age

	Black Females, 1993				White Females, 1993			
x	l_x	${}_nL_x$	T_x	e_x	l_x	${}_nL_x$	T_x	e_x
0-1	100,000	98,702	7,418,554	74.2	100,000	99,464	7,977,549	79.8
1-4	98,552	393,515	7,319,853	74.3	99,398	397,269	7,878,085	79.3
5-9	98,279	490,998	6,926,338	70.5	99,265	496,109	7,480,816	75.4
…								
85+	30,497	232,215	232,214	7.6	42,768	296,729	296,729	6.9

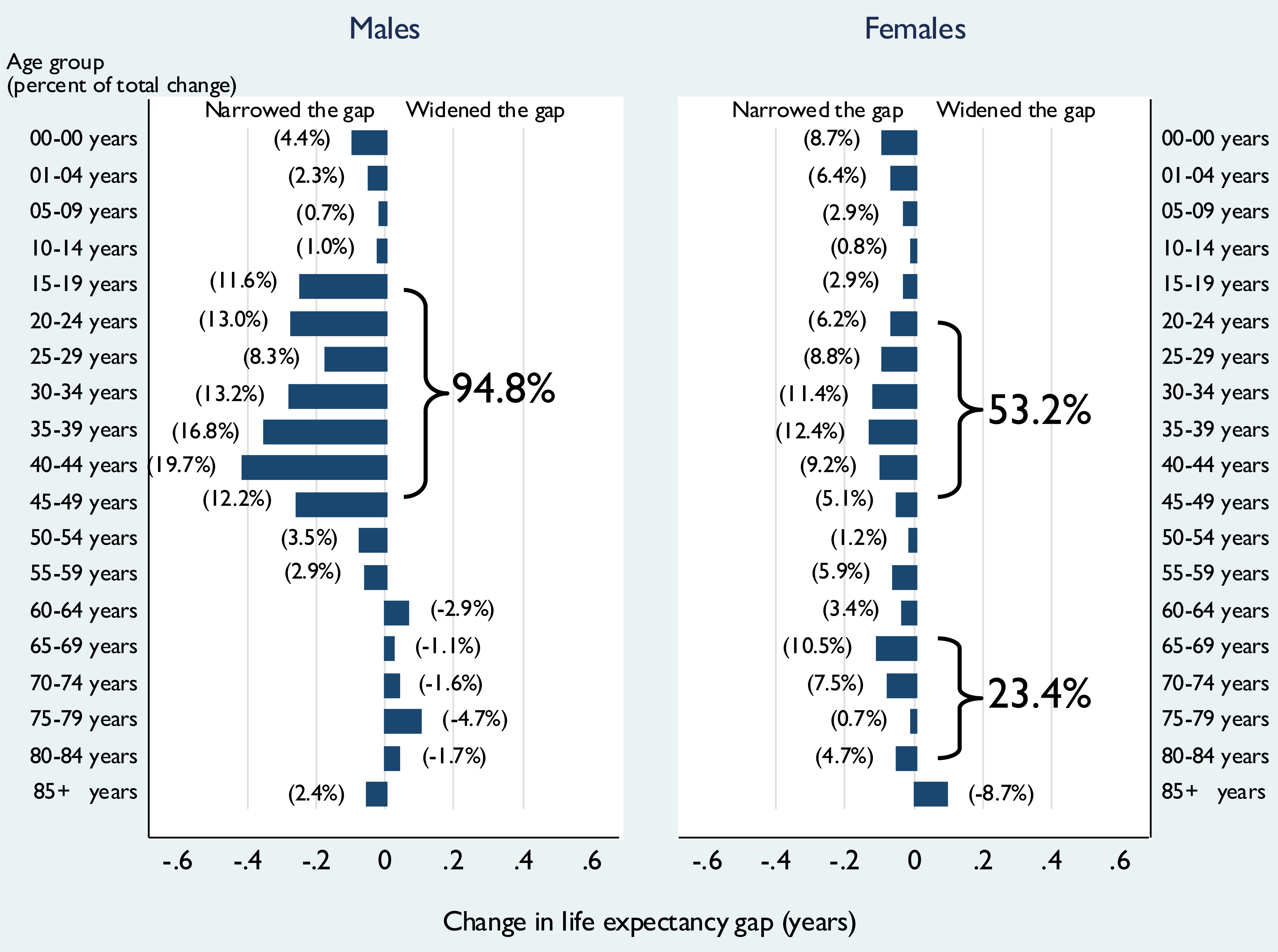
$$
{}_4\Delta_1 = \left[\frac{98552}{100000}\times\left(\frac{397269}{99398}-\frac{393515}{98552}\right)\right] + \left[\frac{7480816}{99265}\times\frac{\frac{98552\times 99265}{99398}-98279}{100000}\right]
$$

$$
= \left[0.986\times(3.997-3.993)\right]+\left[75.36\times\frac{98420.13-98279}{100000}\right]
$$

$$
= 0.004 + 0.106
$$

$$
= 0.110 \text{ years}
$$

(Arriaga, 1984)


Males
Females
Age group
(percent of total change)
Narrowed the gap
Widened the gap
00-00 years (4.4%) (8.7%)
01-04 years (2.3%) (6.4%)
05-09 years (0.7%) (2.9%)
10-14 years (1.0%) (0.8%)
15-19 years (11.6%) (2.9%)
20-24 years (13.0%) (6.2%)
25-29 years (8.3%) (8.8%)
30-34 years (13.2%) (11.4%)
35-39 years (16.8%) (12.4%)
40-44 years (19.7%) (9.2%)
45-49 years (12.2%) (5.1%)
50-54 years (3.5%) (1.2%)
55-59 years (2.9%) (5.9%)
60-64 years (-2.9%) (3.4%)
65-69 years (-1.1%) (10.5%)
70-74 years (-1.6%) (7.5%)
75-79 years (-4.7%) (0.7%)
80-84 years (-1.7%) (4.7%)
85+ years (2.4%) (-8.7%)
94.8%
53.2%
23.4%
-.6 -.4 -.2 0 .2 .4 .6
Change in life expectancy gap (years)

Decomposing by Age and Cause of Death

Percentage of deths by cause, 1993

Age	Total Age Effect	White				Black		
x	${}_n\Delta_x$	Perinatal	CHD	Injuries	.....	Perinatal	CHD	Injuries
00-01	0.68	41.4	2.0	2.9		53.8	2.2	2.3
01-04	0.11	0.9	3.9	35.3		1.2	5.1	31.0
05-09	0.05	0.2	4.2	33.3		0.4	4.5	40.8
...								
75-79	0.29	0.0	34.3	1.6		0.0	37.4	1.4
80-84	0.16	0.0	38.8	1.6		0.0	40.0	1.4
85 +	-0.21	0.0	44.6	1.6		0.0	43.3	1.3
Total	5.59	← Black-White difference in life expectancy at birth, 1993						

“Decomposing” Life Expectancy Differences by Cause

Difference in share of deaths due to cause *i*

$$ {}_n\Delta_x^i = {}_n\Delta_x \times \frac{\left({}_np_x^{i,White} \times {}_nr_x^{White}\right) - \left({}_np_x^{i,Black} \times {}_nr_x^{Black}\right)}{{}_nr_x^{White} - {}_nr_x^{Black}} $$

Total contribution of age group

Difference in age-specific mortality for all causes

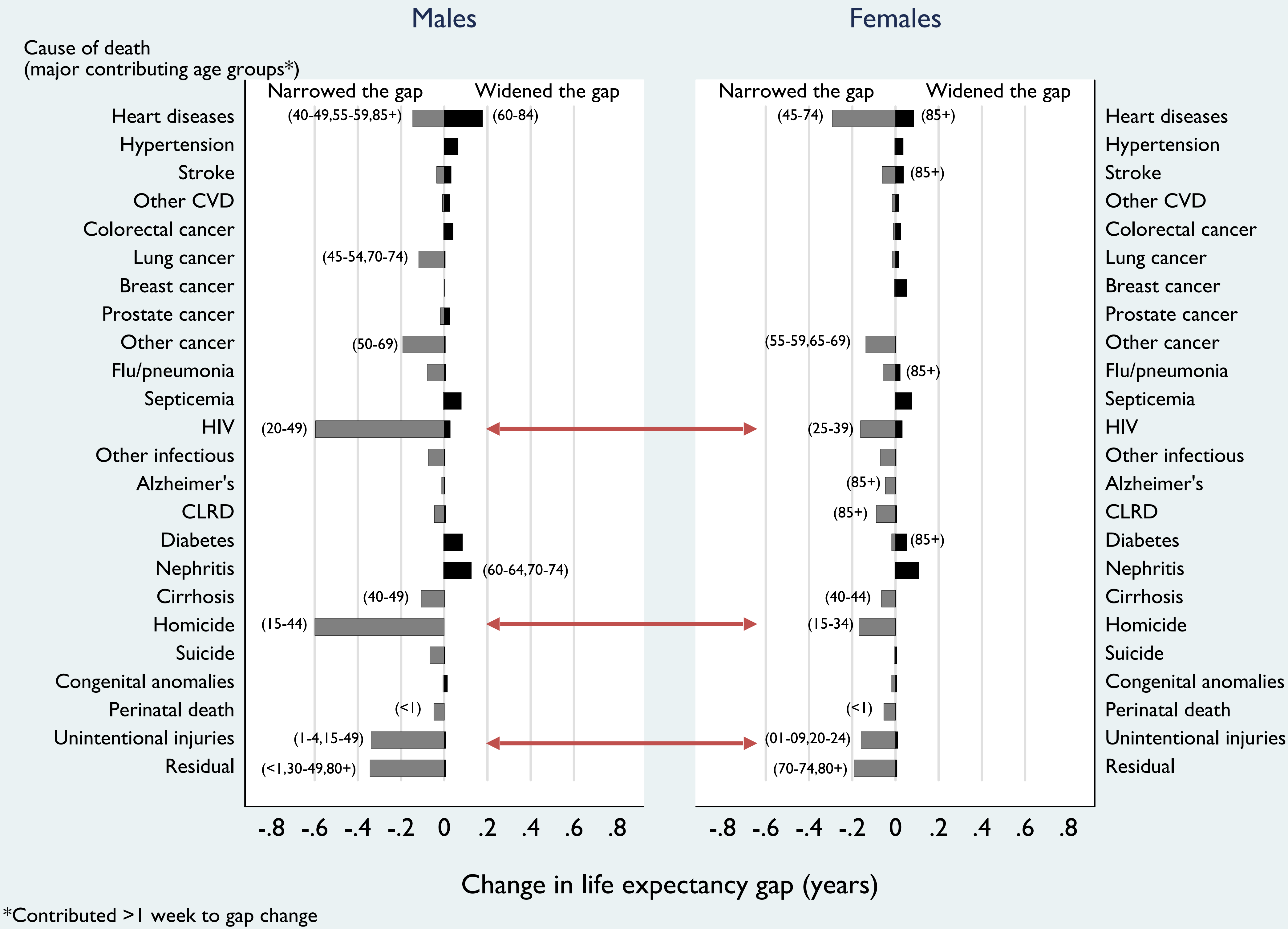
(Arriaga, 1989)

Decomposing by Age and Cause of Death

Age	Total Age Effect	Percentage of deaths by cause, 1993						
		White				Black		
x	${}_n\Delta_x$	Perinatal	CHD	Injuries	.....	Perinatal	CHD	Injuries
00-01	0.68	41.4	2.0	2.9		53.8	2.2	2.3

$$
\begin{aligned}
{}_1\Delta_0^{Perinatal} &= 0.68 \times \frac{\left(0.414^{Perinatal}_{white} \times 6.05^{Rate}_{white}\right) - \left(0.538^{Perinatal} \times 14.67^{Rate}_{black}\right)}{6.05^{Rate}_{white} - 14.67^{Rate}_{black}} \\
&= 0.68 \times \left[(2.51 - 7.90)/(-8.62)\right] \\
&= 0.68 \times \left[(-5.39)/(-8.62)\right] \\
&= 0.68 \times 0.63 \\
&= 0.42
\end{aligned}
$$

Thus, 63% (0.42/0.68) of the total contribution among infants is due to the difference between blacks and whites in rates of perinatal death



Males
Females
Cause of death
(major contributing age groups*)
Narrowed the gap
Widened the gap
Heart diseases
Hypertension
Stroke
Other CVD
Colorectal cancer
Lung cancer
Breast cancer
Prostate cancer
Other cancer
Flu/pneumonia
Septicemia
HIV
Other infectious
Alzheimer's
CLRD
Diabetes
Nephritis
Cirrhosis
Homicide
Suicide
Congenital anomalies
Perinatal death
Unintentional injuries
Residual
(40-49,55-59,85+)
(60-84)
(45-54,70-74)
(50-69)
(20-49)
(60-64,70-74)
(40-49)
(15-44)
(<1)
(1-4,15-49)
(<1,30-49,80+)
(45-74)
(85+)
(85+)
(55-59,65-69)
(85+)
(25-39)
(85+)
(85+)
(85+)
(40-44)
(15-34)
(<1)
(01-09,20-24)
(70-74,80+)
-.8 -.6 -.4 -.2 0 .2 .4 .6 .8
Change in life expectancy gap (years)
*Contributed >1 week to gap change

Determinants of Socioeconomic Inequalities: Decomposition of the Concentration Index

Decomposition: Moving from Description to Explanation

- The "substance" we want to decompose is health inequality—a difference in health between social groups.

- Ultimately, we want to know <u>why</u> health inequalities are changing over time, or why they differ between populations
  - Risk factors?
  - Demographic composition?
  - Social conditions?

- Unpacking the 'components' of health inequality is an opportunity to better integrate the monitoring of health inequalities with the etiology of health inequalities.

- These techniques often involve various kinds of 'counterfactual' scenarios

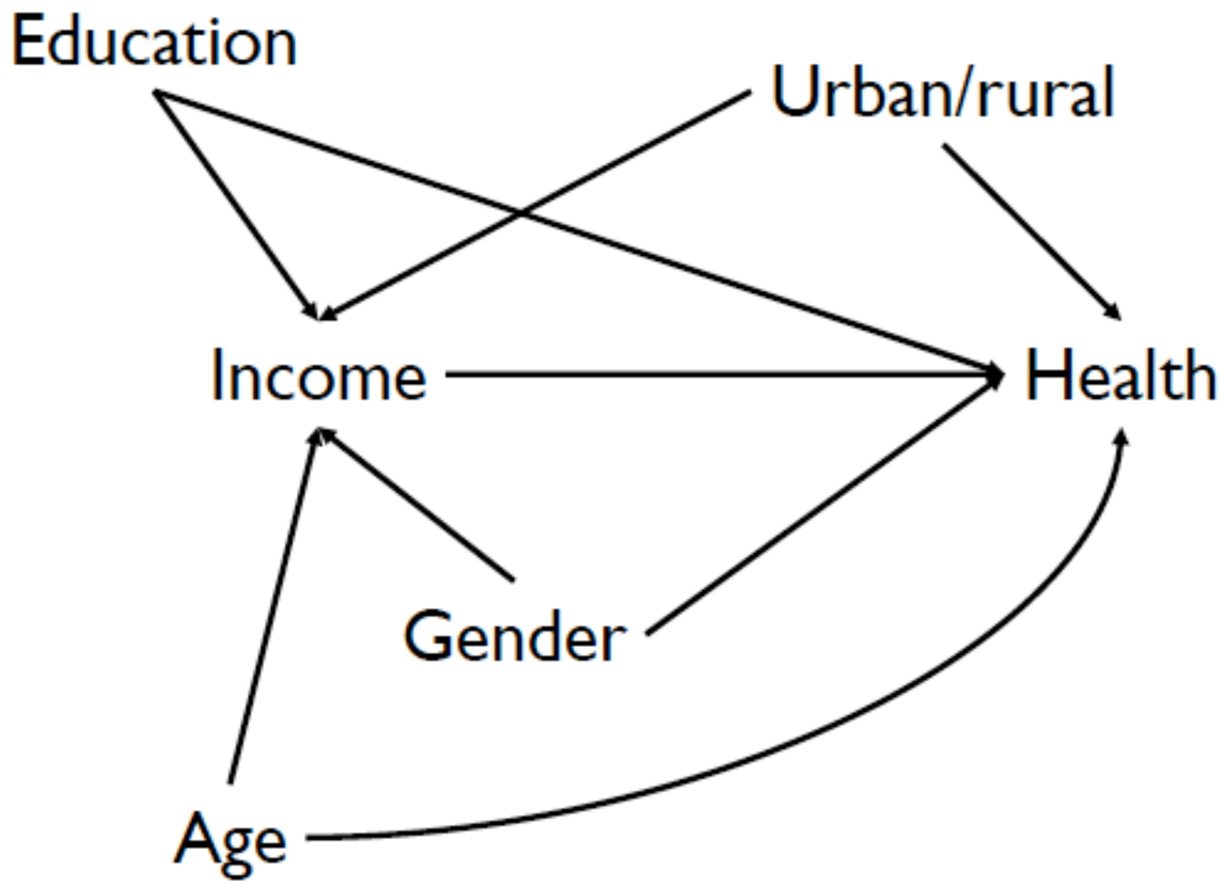
The usual approach

- Conventional methods for “explaining” effects of social exposures
  - Estimate crude or demographic-adjusted effect (logit, hazard)
  - Add “conventional” risk factors (physiological, behavioural)
  - Add “novel” risk factors (flavour-of-the-week)
  - Interpret accordingly

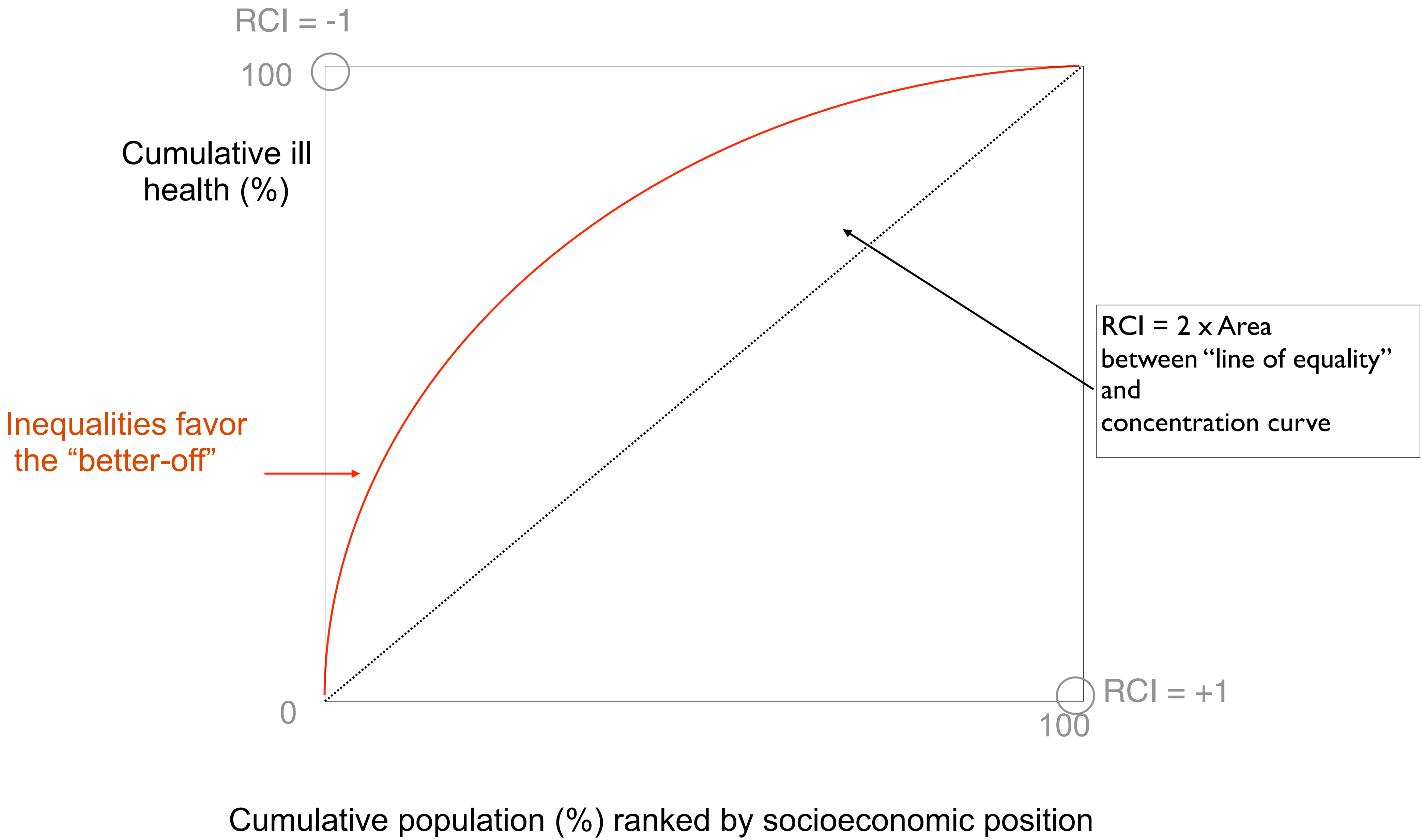
- Limitations of conventional approach
  - Often fail to consider entire socioeconomic distribution (typically low vs. high only) in the context of “explanation”
  - Often ignore absolute risk
  - Typically do not provide estimates of the specific contributions of other factors to the “explained” proportion

We want to understanding the components of this:

By estimating something like this:



Relative Concentration Index



Formula for Calculating the Relative Concentration Index

One way of writing the Relative Concentration Index* is

$$RCI = \frac{2}{n\mu}\sum_{i=1}^{n} y_i R_i - 1$$

Where μ is the mean of y_i (e.g., smoking status), R_i is the fractional rank of the ith person in the socioeconomic (e.g., income) distribution

The Absolute Concentration Index simply multiplies RCI by the mean smoking rate:

$$ACI = \mu RCI$$

*(Kakwani et al. 1997)

Decomposition of the Concentration Index

On decomposing the causes of health sector inequalities with an application to malnutrition inequalities in Vietnam

Adam Wagstaff[a,b,*], Eddy van Doorslaer[c], Naoko Watanabe[a]

Journal of Econometrics 112 (2003) 207–223

- Recall we can write the *RCI* as: $$RCI = \frac{2}{n\mu}\sum_{i=1}^{n} y_i R_i - 1$$
- Suppose that one can write a regression equation expressing the health outcome of interest (y_i) as a function of several (k_i) determinants (e.g., age, gender, urban/rural status):

$$y_i = \alpha + \sum_k \beta_k x_{ki} + \varepsilon_i$$

Decomposition of the Concentration Index

- Given that RCI is a function of yi and socioeconomic rank, one can then re-express the concentration index RCI for y as:

$$RCI = \sum_k \left(\beta_k \bar{x}_k / \mu\right) RCI_k + gRCI_\varepsilon / \mu$$

- Where μ is the mean of y, $\bar{x}_k$ is the mean of x_k, β_k is the regression coefficient for x_k, and RCI_k is the concentration index for x_k.

- The basic ideas here is to determine how much of the overall inequality is due to other factors that are both differentially distributed by income and affect smoking

“Explained” and “unexplained” components

This equation results in 2 components of socioeconomic inequality:

$$RCI = \sum_k (\beta_k \bar{x}_k / \mu) RCI_k + gRCI_\varepsilon / \mu$$

One part that is due to the association between income and other factors that predict health

The other part is ‘unexplained’, i.e., inequality that cannot be explained by systematic variation across income groups in the determinants of health.

2 types of "explained" components

The influence of other determinants depends on 2 things:

$$RCI = \sum_k \overbrace{\left(\beta_k \bar{x}_k / \mu\right) RCI_k}^{} + gRCI_\varepsilon / \mu$$

RCI_k : the strength of the relationship between each factor and income (the concentration index for each determinant)

$\beta_k \bar{x}_k / \mu$: the strength of the relationship between each factor and health, and its prevalence in the population (also referred to as the 'elasticity' of each factor with respect to health).

Procedure for decomposing the Concentration Index

1. Estimate a regression equation predicting y ('health') from its determinants (β_k)

$$y_i = \alpha + \sum_k \beta_k x_{ki} + \varepsilon_i$$

2. Calculate the mean of y (μ) and of each of the determinants (e.g., education, age)

3. Calculate the Concentration Index for the health variable (C) each determinant in the equation predicting health (Ck).

   - That is, use each determinant as the "outcome" and estimate a CI for age, CI for education, etc.

Procedure for decomposing the Concentration Index

4. Calculate the absolute contribution of each determinant by multiplying its ‘elasticity’ by its C:

$$(\beta_k \bar{x}_k / \mu) RCI_k$$

5. Calculate the percentage contribution of each determinant:

$$[(\beta_k \bar{x}_k / \mu) RCI_k] / RCI$$

A few examples

Decomposing socioeconomic inequality in infant mortality in Iran

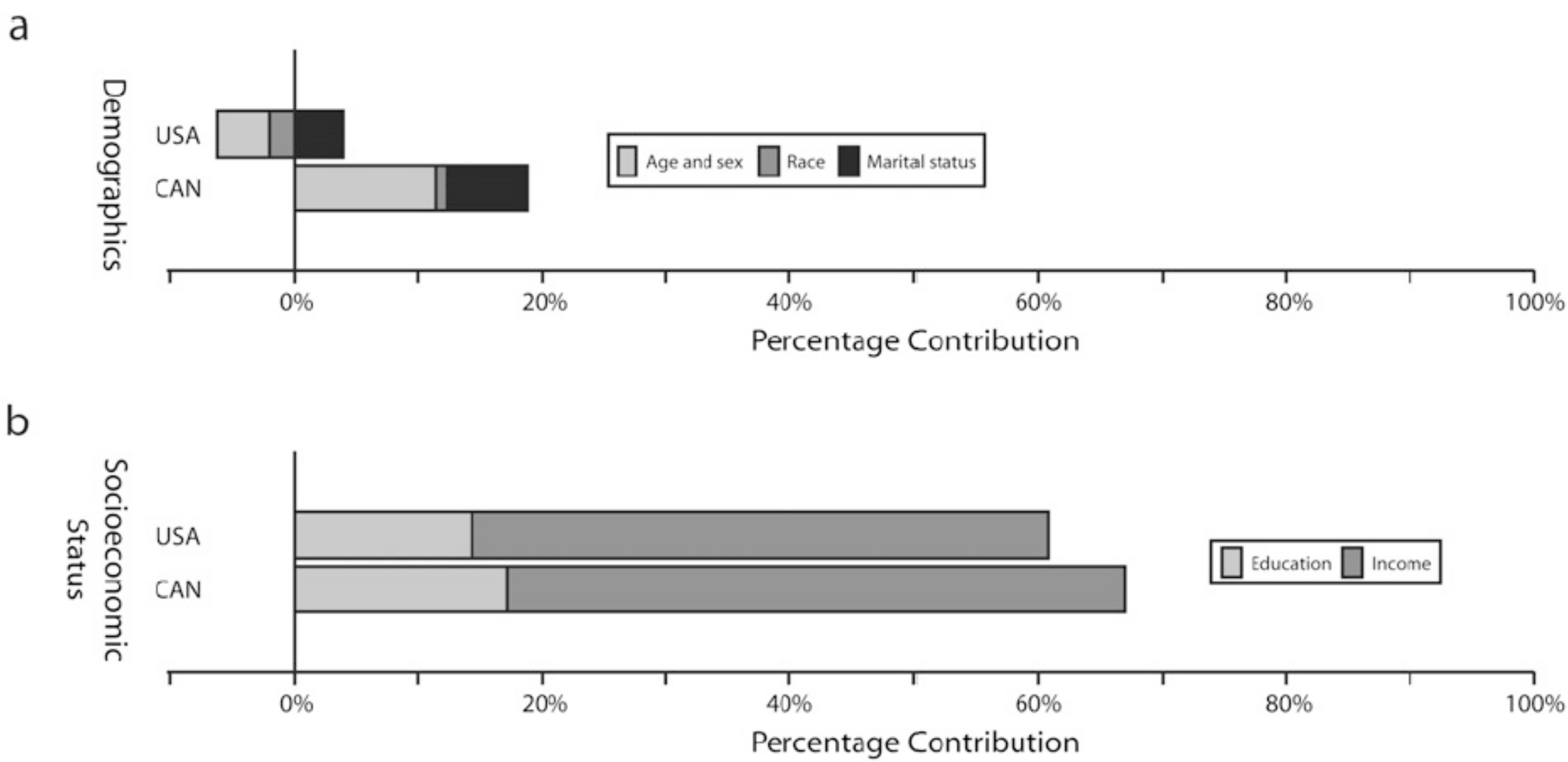
Ahmad Reza Hosseinpoor,[1,*] Eddy Van Doorslaer,[2] Niko Speybroeck,[1] Mohsen Naghavi,[3]
Kazem Mohammad,[4] Reza Majdzadeh,[4] Bahram Delavar,[3] Hamidreza Jamshidi[3] and Jeanette Vega[1]

Overall Concentration index for economic status and infant mortality = 0.0413

Determinant	Beta coef.	Mean of x	Ck	Contrib to C	% of C
History of mother's stillbirth	0.5643	0.0650	-0.1001	.0010	2.5
History of mother's abortion	0.1313	0.2146	0.0396	-0.0003	-0.8
Risky birth interval	0.8028	0.1664	-0.1426	0.0054	13.0
Low economic status	0.2287	0.3634	-0.6366	0.0150	36.2
Mother's illiteracy	0.3088	0.3524	-0.2803	0.0086	20.9
Having a hygienic toilet	-0.1700	0.2916	0.3503	0.0049	11.9
Rural residency	0.1706	0.4470	-0.2663	0.0057	13.9
Total				0.0413	100.0

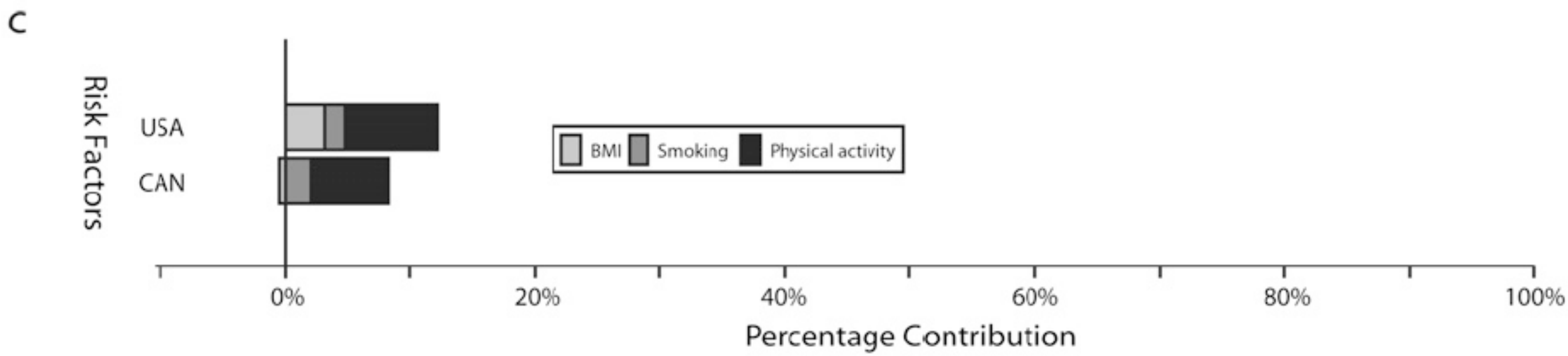
Income-Related Health Inequalities in Canada and th United States: A Decomposition Analysis

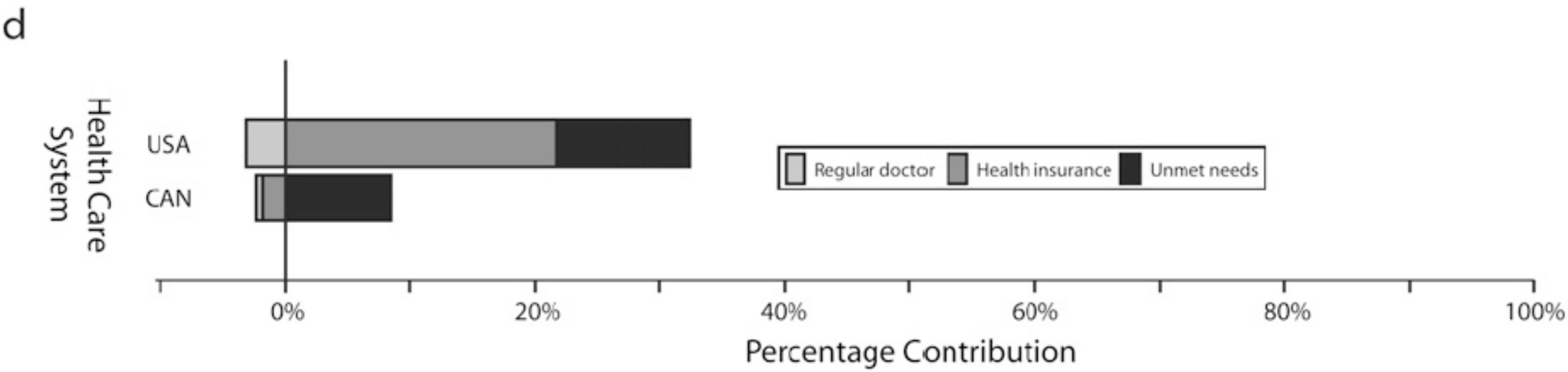
Kimberlyn M. McGrail, PhD, Eddy van Doorslaer, PhD, Nancy A. Ross, PhD, and Claudia Sanmartin, PhD



Income-Related Health Inequalities in Canada and th United States: A Decomposition Analysis

Kimberlyn M. McGrail, PhD, Eddy van Doorslaer, PhD, Nancy A. Ross, PhD, and Claudia Sanmartin, PhD

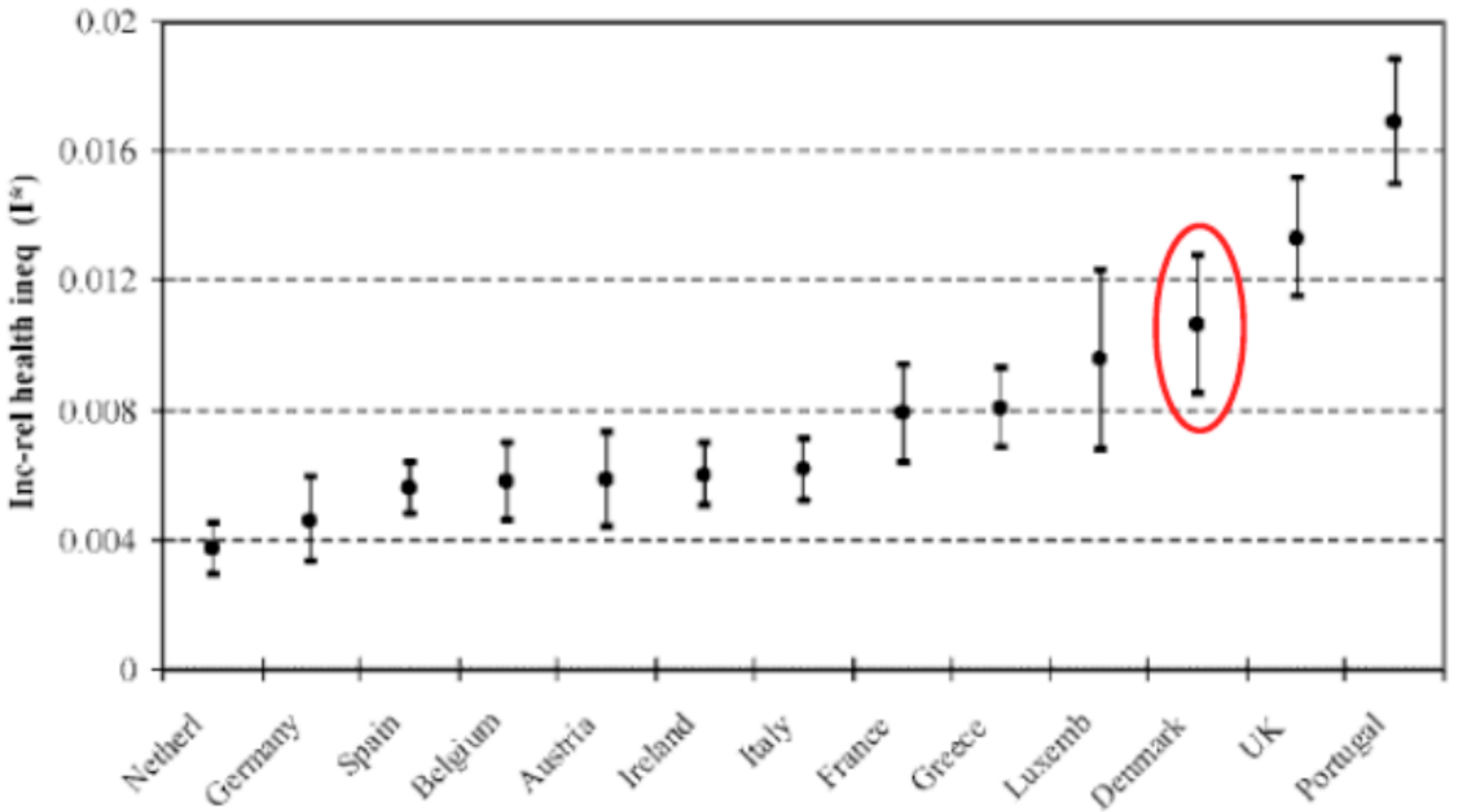




HEALTH ECONOMICS
Health Econ. **13**: 609–628 (2004)
Published online 23 June 2004 in Wiley InterScience (www.interscience.wiley.com). **DOI**:10.1002/hec.918

Explaining the differences in income-related health inequalities across European countries

Age-sex standardized income-related health inequality in self-assessed health across European countries



"The peculiar Danish result does not arise because Denmark has the lowest income inequality in the EU, but because of the complete lack of any (partial) linear association between household income and adult health."

Table 4. Health inequality contributions of regressors per country (in % of HUI conc index, and with bootstrapped *t*-values)

	Germany		Denmark		Netherlands		Belgium		Luxembourg		France		UK	
	CI contr	*t*-val	CI contr	*t*-val	CI contr	*t*-val	CI contr	*t*-val	CI contr	*t*-val	CI contr	*t*-val	CI contr	*t*-val
C (HUI pred)	0.00434		0.00938		0.00337		0.00710		0.01036		0.00745		0.01286	
$I^* = C - C^*$	0.00461	7.05	0.01062	9.85	0.00372	9.22	0.00579	9.37	0.00955	6.80	0.00788	10.25	0.01332	14.39
Ln (Inc)	**36.7%**	2.68	0.9%	0.10	31.6%	2.74	33.9%	4.49	45.8%	3.80	36.1%	3.30	25.8%	4.44
M30–44	**−5.0%**	−3.90	−4.6%	−4.29	−2.4%	−2.42	−4.8%	−3.82	−2.4%	−1.87	−3.0%	−2.94	−2.4%	−3.42
M45–59	**−21.0%**	−7.10	−8.8%	−5.47	−14.8%	−6.39	−4.5%	−3.50	−1.7%	−0.97	−13.0%	−7.50	−5.0%	−5.05
M60–69	0.3%	0.47	−0.3%	−0.29	−0.8%	−1.13	2.1%	1.79	0.0%	−0.07	−1.3%	−1.26	0.0%	0.12
M70 +	0.2%	0.12	0.3%	0.13	1.0%	1.21	5.7%	3.85	1.5%	0.91	1.5%	1.06	0.6%	0.54
F16–29	**1.6%**	2.15	0.3%	0.56	1.2%	1.39	0.5%	1.01	0.1%	0.24	1.6%	2.74	0.1%	0.27
F30–44	0.5%	0.82	−4.3%	−4.09	2.1%	1.94	−1.1%	−1.22	0.5%	0.40	0.5%	0.58	−0.2%	−0.86
F45–59	**−11.6%**	−5.18	−6.4%	−4.66	−13.1%	−5.52	−3.5%	−2.42	−0.7%	−0.28	−11.5%	−6.71	−2.8%	−3.80
F60–69	4.5%	2.84	1.2%	0.72	2.1%	1.81	5.6%	3.10	1.6%	0.95	1.2%	1.11	0.0%	−0.05
F70 +	24.3%	5.42	9.2%	2.35	14.3%	4.35	18.4%	5.73	8.9%	2.66	18.4%	6.64	6.0%	2.60
Second educ	0.0%	0.00	−1.0%	−1.12	−0.6%	−0.74	0.5%	1.27	8.4%	3.97	2.7%	3.22	1.6%	2.98
Higher educ	15.6%	4.67	18.1%	5.81	22.1%	4.89	10.9%	3.99	14.7%	3.31	21.1%	6.01	14.2%	7.30
Part-time empl	0.0%	−0.11	0.1%	0.34	−0.8%	−1.59	4.5%	2.33			0.1%	0.47	0.0%	−0.19
Self-employed	0.0%	0.02	−0.2%	−0.55	−0.1%	−0.14	−0.4%	−1.20	0.1%	0.21	−0.3%	−0.90	−0.8%	−1.25
Student	−2.5%	−3.01	−1.7%	−2.46	−5.5%	−2.65	−0.5%	−1.28	−1.8%	−1.88	−1.3%	−2.40	0.3%	1.29

Int J Public Health
DOI 10.1007/s00038-010-0224-6

ORIGINAL ARTICLE

Decomposing income-related inequality in cervical screening in 67 countries

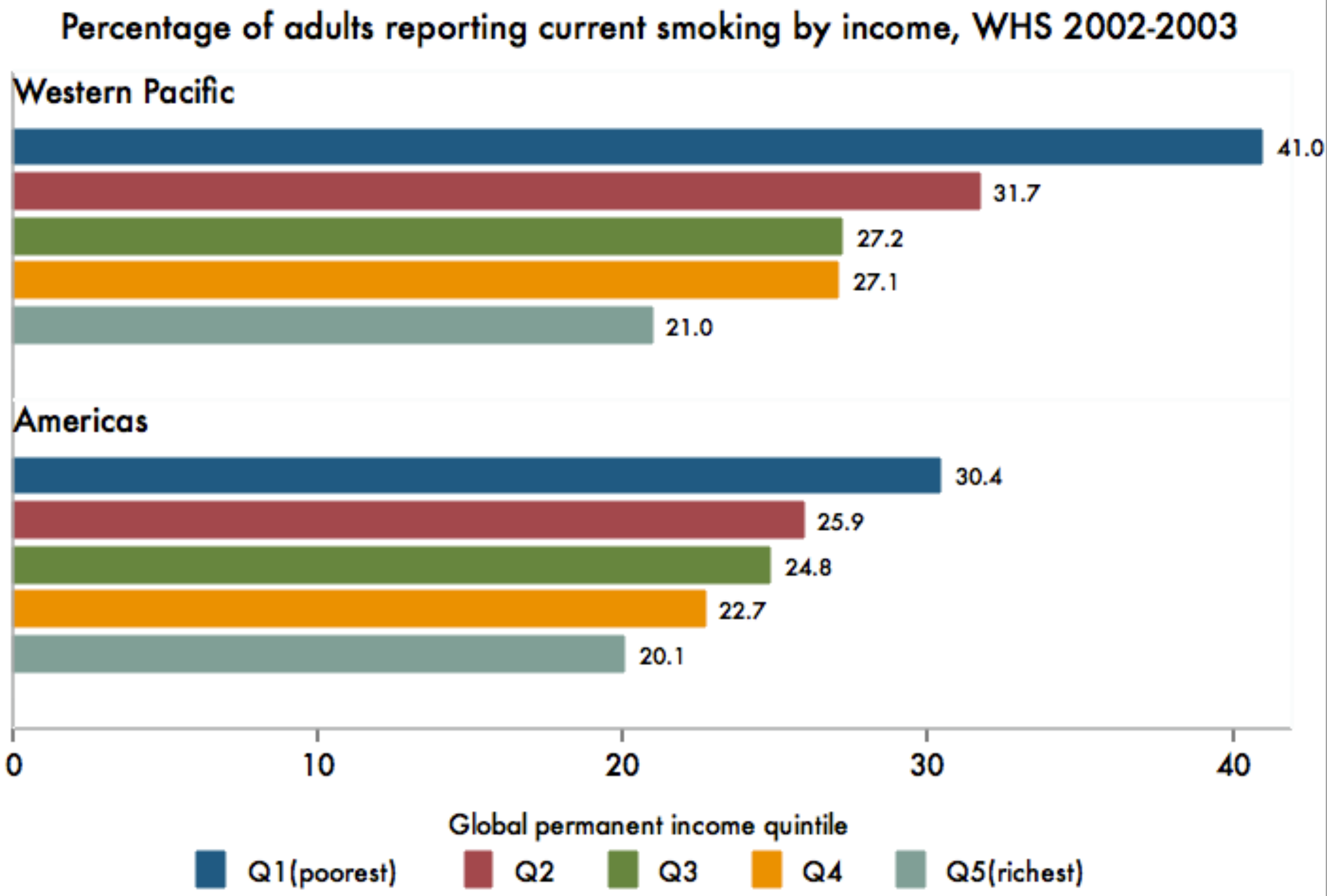
**Brittany McKinnon · Sam Harper ·
Spencer Moore**

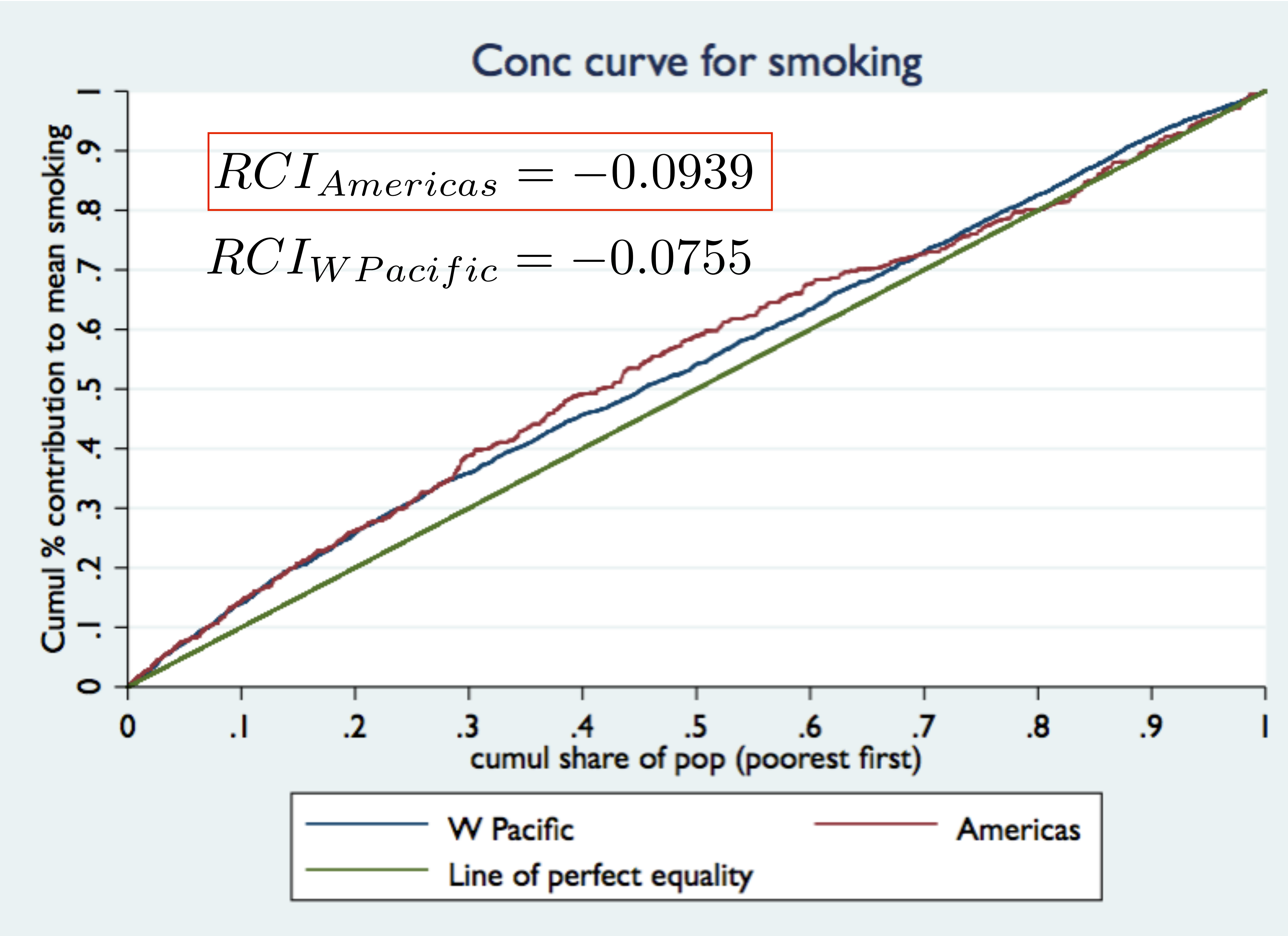
Contribution of education to income-related inequality in screening was highly variable across countries

Table 4 Percentage contribution of determinants to income-related inequality in cervical screening, World Health Survey 2002–2003

WHO region	Country	Age	Income	Urban	Marital Status	Education	Recent health care[a]	Unexplained
Africa	Chad	0.1	47.2	5.2	−0.7	−2.1	5.8	58.8
	Côte d'Ivoire	48.1	−0.7	15.8	−14.0	42.6	2.9	12.8
	Ethiopia	−0.6	34.2	9.8	1.4	6.0	2.6	44.4
	Ghana	−3.1	79.4	−6.4	−4.7	12.2	3.2	20.6
	Kenya	0.0	61.8	2.3	−4.3	15.3	−0.7	29.8
	Mali	−1.5	32.5	26.1	0.4	0.0	10.9	31.6
	Mauritania	2.0	11.9	18.0	−0.4	−6.4	5.8	42.9
	Mauritius	3.5	87.3	7.3	4.3	−3.0	−6.7	18.1
	Namibia	3.4	59.9	16.2	2.5	4.9	4.2	8.8
	Senegal	−8.9	83.9	2.7	−22.2	50.6	5.9	−20.3
	South Africa	2.4	46.2	14.3	7.2	33.0	−0.7	−2.7
	Swaziland	0.3	65.3	−2.5	0.0	15.7	0.9	20.2
	Zambia	19.4	15.2	26.3	1.2	9.1	0.0	31.1
Americas	Brazil	−2.4	64.5	−2.1	4.5	39.9	4.5	−8.9

Example: Decomposing Socioeconomic Inequality in Current Smoking


Percentage of adults reporting current smoking by income, WHS 2002-2003
Western Pacific
41.0
31.7
27.2
27.1
21.0
Americas
30.4
25.9
24.8
22.7
20.1
0
10
20
30
40
Global permanent income quintile
Q1(poorest)
Q2
Q3
Q4
Q5(richest)



Conc curve for smoking
$RCI_{Americas} = -0.0939$
$RCI_{WPacific} = -0.0755$
Cumul % contribution to mean smoking
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1
cumul share of pop (poorest first)
W Pacific
Americas
Line of perfect equality

Estimation for a specific factor: Education

Recall the decomposition formula:

$$RCI = \sum_{k=1}^{K} (\beta_k \bar{x}_k/\mu)\, RCI_k + gRCI_\varepsilon/\mu$$

Estimated β coeff on education (**logit scale**): -.0389 (OR = 0.96)

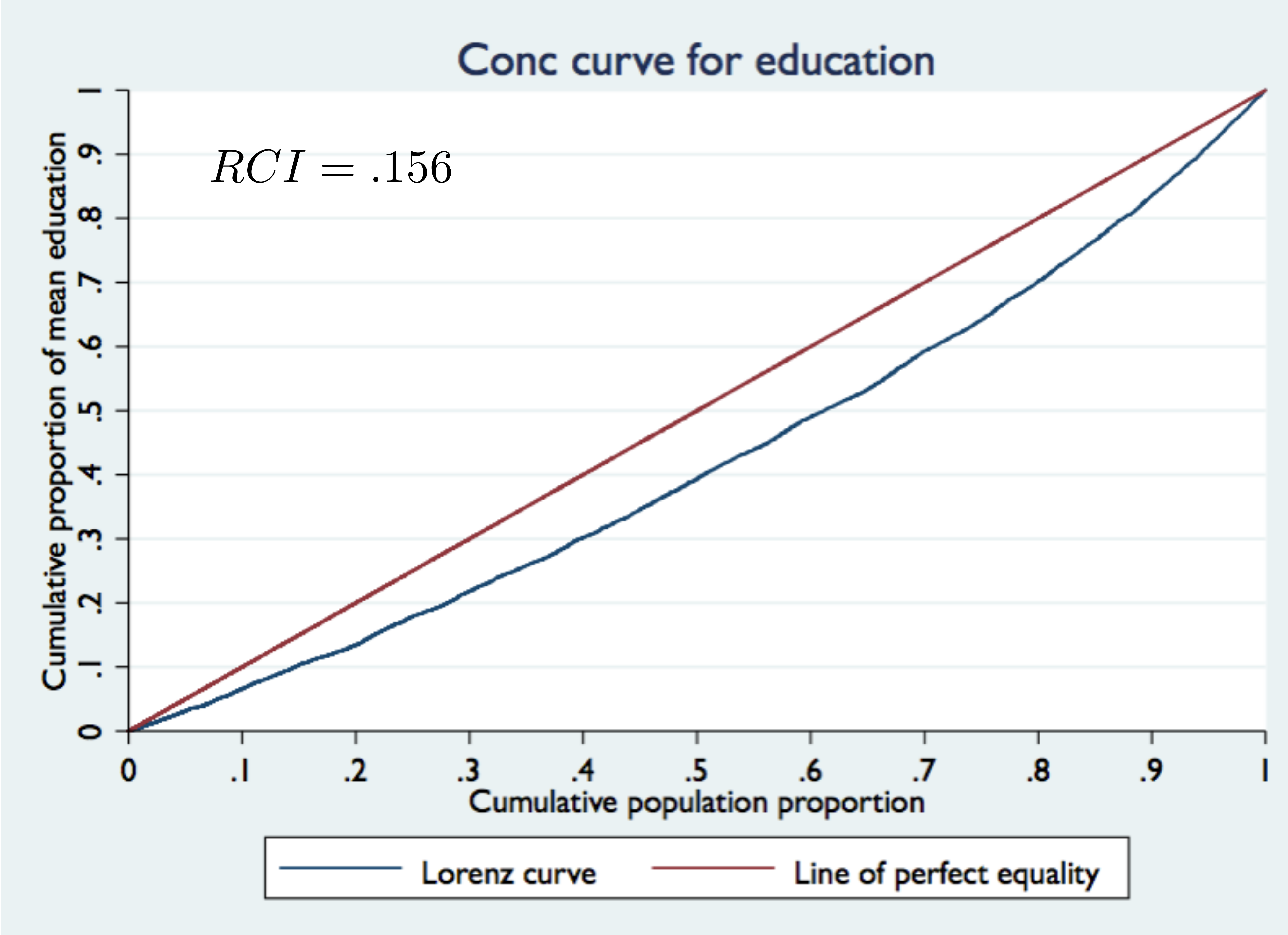
Marginal effect on **probability/absolute** scale: -.0051 (0.5 pct points)

Mean education: 8.9 yrs

Mean smoking rate: 17.5%

With these parameters, the elasticity of smoking with respect to education is:
(-.0051 * 8.9 / .175) = -.2582

Interpretation: a 1% increase in education decreases smoking by 26% (not percentage points!). What about the RCI for education?


Conc curve for education
$RCI = .156$
Cumulative proportion of mean education
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1
Cumulative population proportion
Lorenz curve
Line of perfect equality

Estimation for a specific factor: Education

Recall the decomposition formula:

$$RCI = \sum_{k=1}^{K} \left(\beta_k \bar{x}_k / \mu\right) RCI_k + gRCI_\varepsilon / \mu$$

So the elasticity of smoking (from the previous slide) with respect to education is (-.0051 * 8.9 / .175) = -.2582

Now we have the RCI for education = 0.156

So now we can calculate the contribution of education as:

Elasticity * RCI_{ed} = -.2582 * .156 = -.04

Thus education accounts for -.04/ -.0939 = 41.6%

Decomposition of Income-Related Inequality in Smoking: Americas region (overall RCI = -0.094)

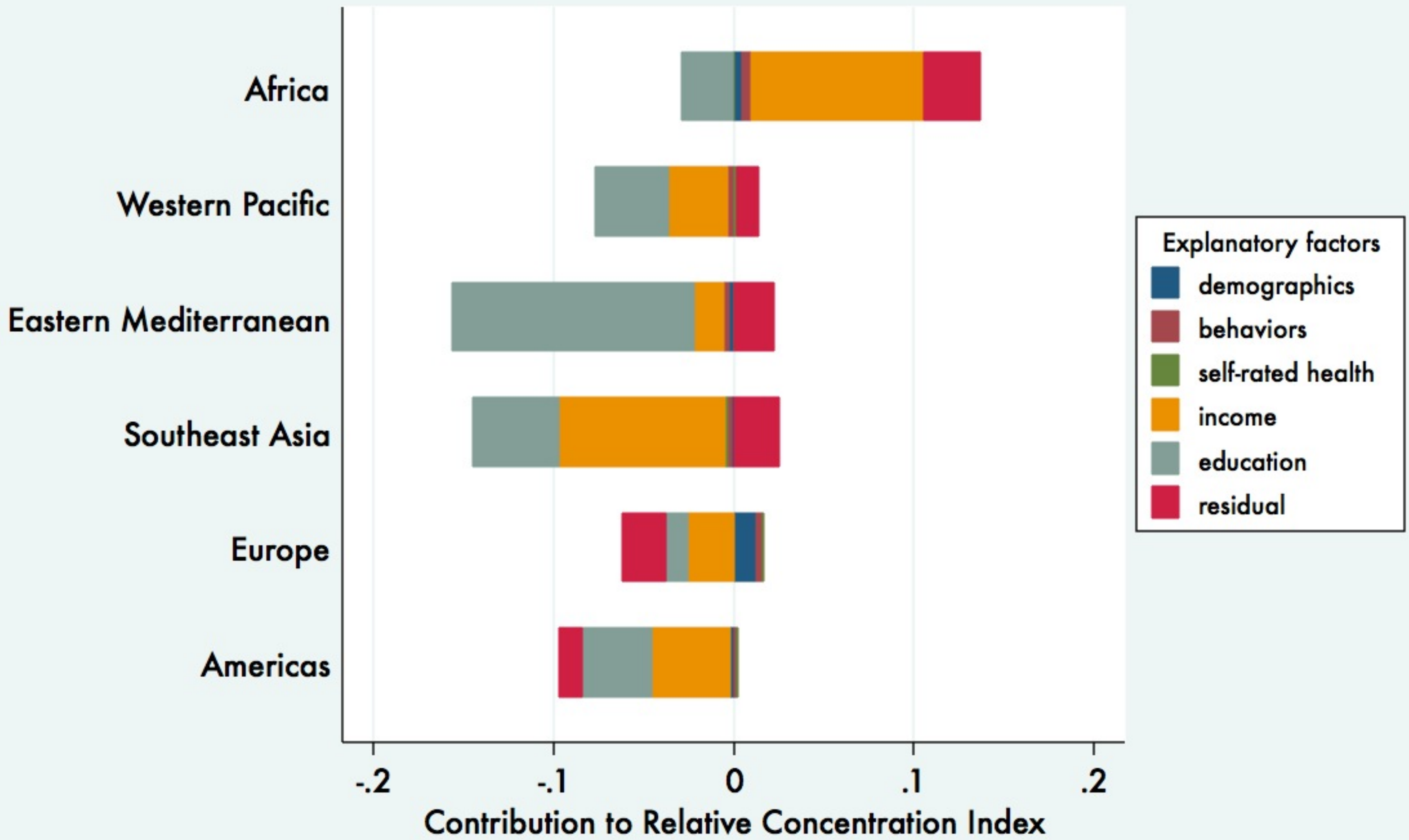
	Elasticity	Rel Conc Index	Contribution	% Contrib
Age	3.695	0.023	0.084	-89.9%
Age^2	-1.981	0.032	-0.064	67.9%
Male	0.197	-0.055	-0.011	11.5%
BMI	-0.834	0.011	-0.009	9.6%
Urban	0.020	0.076	0.002	-1.6%
Single	0.078	-0.036	-0.003	3.0%
Divorced/Widowed	0.161	-0.120	-0.019	20.7%
Low Phys Activity	0.057	0.069	0.004	-4.2%
Mod Phys Activity	-0.023	0.025	-0.001	0.6%
Low Alcohol Consumption	0.131	0.123	0.016	-17.1%
Mod/Hi Alcohol Consumption	0.019	0.081	0.002	-1.6%
Low Fruit/Veg Consumption	0.029	-0.066	-0.002	2.0%
Self-Reported Health Good	-0.001	0.040	0.000	0.1%
Self-Reported Health Moderate	-0.043	-0.079	0.003	-3.6%
Self-Reported Health Bad/Very Bad	0.004	-0.208	-0.001	0.9%
Education	-0.250	0.156	-0.039	41.6%
Permanent Income	-0.809	0.054	-0.044	46.4%
Residual			-0.013	

Contrasting components of income-related inequality

	Elasticity	RCI	Contribution	% Contribution
Western Pacific				
Income	-0.51	0.065	-0.033	43.7%
Urbanicity	0.06	0.252	0.016	-20.8%
Education	-0.43	0.096	-0.041	54.5%
Americas				
Income	-0.81	0.054	-0.044	46.4%
Urbanicity	0.02	0.076	0.002	-1.6%
Education	-0.25	0.156	-0.039	41.6%

What does this imply for interventions to reduce income-related inequalities?

Aggregate contribution of explanatory factors to income-related inequalities in current smoking by WHO Region



Caveats for decomposition of the Concentration Index

- Decomposition results will be sensitive to the choice of determinants included (i.e., how well-specified the model is for predicting y).

- The regression equations are predictive and not causal models.

- Main utility is not in estimating the potential impact on y of changing the distribution of socioeconomic position, but in indicating the potential role that other factors may play in generating socioeconomic inequalities in health.

Determinants of Mean Differences: Blinder-Oaxaca Decomposition

Idea

- The core idea is to explain the distribution of the outcome variable in question by a set of factors that vary systematically with exposure status.
- Thus, we want to know, on average, why the mean level of health or disease differs between exposed and unexposed groups.
- Since, for most health outcomes there are multiple determinants, we may want to know which of these determinants plays more or less important roles in explaining the difference in average outcomes.
- “Unpacking” or “decomposing” difference.

Brief note on interpretation

- Decomposition methods are based on regression analyses, and thus all of the usual caveats about good specification apply

- If regressions are purely descriptive, they reveal the associations that characterize the health inequality
  - Then inequality is explained in a statistical sense but implications for policies to reduce inequality are limited

- If data allow identification of causal effects, then the factors that generate the inequality are identified
  - Then one can (potentially) draw conclusions about how policies would impact on inequality

O'Donnell et al. (2008)

Blinder-Oaxaca: Basic idea

Two potential sources of mean differences in outcomes

1. Differences in the prevalence of determinants of outcome

2. Differences in the effect of a given determinant on the outcome (i.e., effect measure modification)

Eur J Health Econ (2011) 12:17–28
DOI 10.1007/s10198-010-0220-z

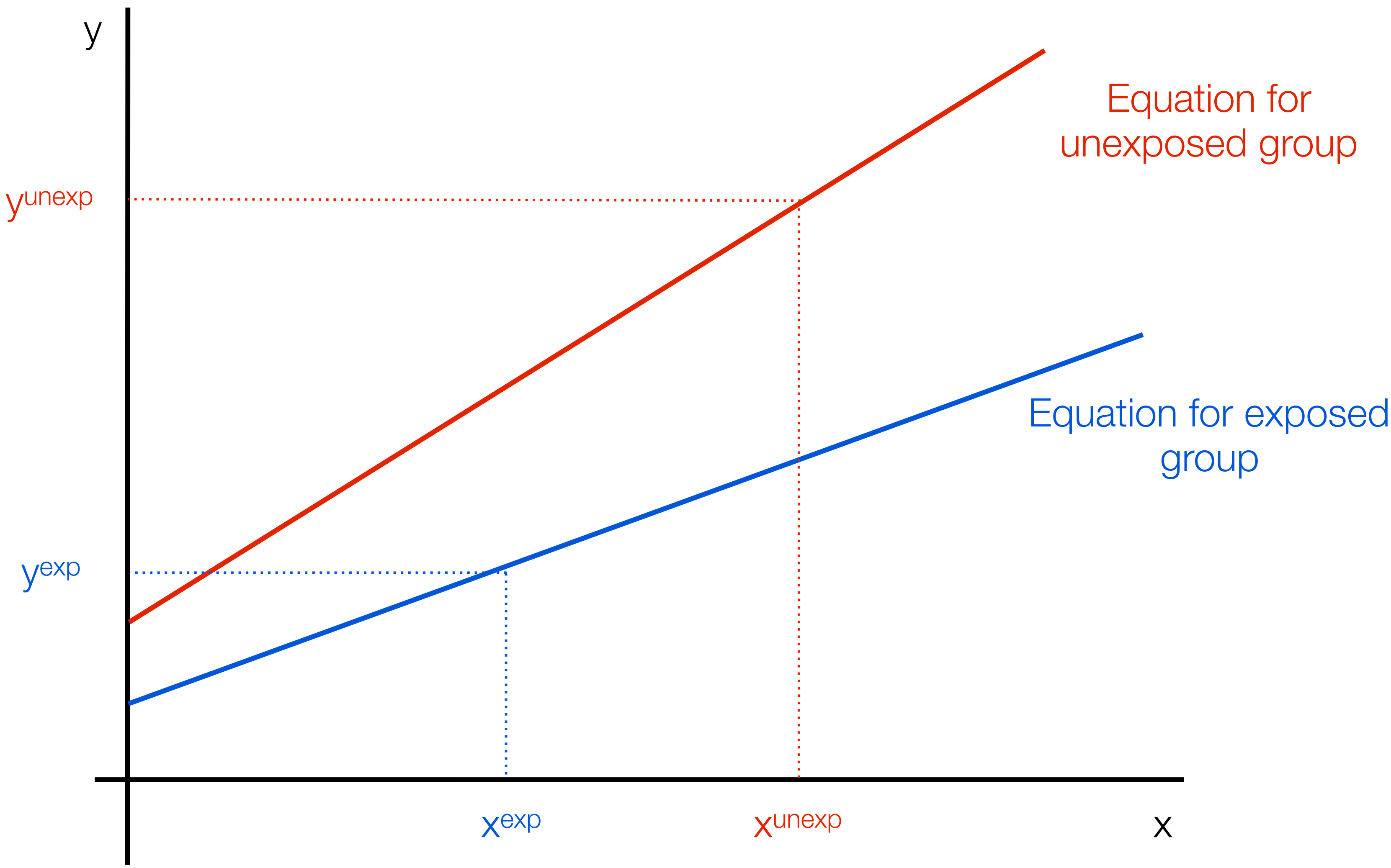
ORIGINAL PAPER

Inequalities in the use of health services between immigrants and the native population in Spain: what is driving the differences?

Dolores Jiménez-Rubio · Cristina Hernández-Quevedo

Abstract In Spain, a growing body of literature has drawn attention to analysing the differences in health and health resource utilisation of immigrants relative to the autochthonous population. The results of these studies generally find substantial variations in health-related patterns between both population groups. In this study, we use the Oaxaca-Blinder decomposition technique to explore to what extent disparities in the probability of using medical care use can be attributed to differences in the determinants of use due to, e.g. a different demographic structure of the immigrant collective, rather than to a different effect of health care use determinants by nationality, holding all other factors equal. Our findings show that unexplained factors associated to immigrant status determine to a great extent disparities in the probability of using hospital, specialist and emergency services of immigrants relative to Spaniards, while individual characteristics, in particular self-reported health and chronic conditions, are much more important in explaining the differences in the probability of using general practitioner services between immigrants and Spaniards.

$$y_i = \left\{ \begin{array}{c} \beta^{\text{exp}} x_i + \varepsilon_i^{\text{exp}} \text{ if exposed} \\ \beta^{\text{unexp}} x_i + \varepsilon_i^{\text{unexp}} \text{ if unexposed} \end{array} \right\}$$



N.B.: 2 ways of expressing the mean difference

The overall gap between exposed and unexposed can be written as a function of differences the respective beta coefficients, evaluated at the mean for each group:

$$y^{exp} - y^{unexp} = \beta^{exp}\bar{x}^{exp} - \beta^{unexp}\bar{x}^{unexp}$$

This way:

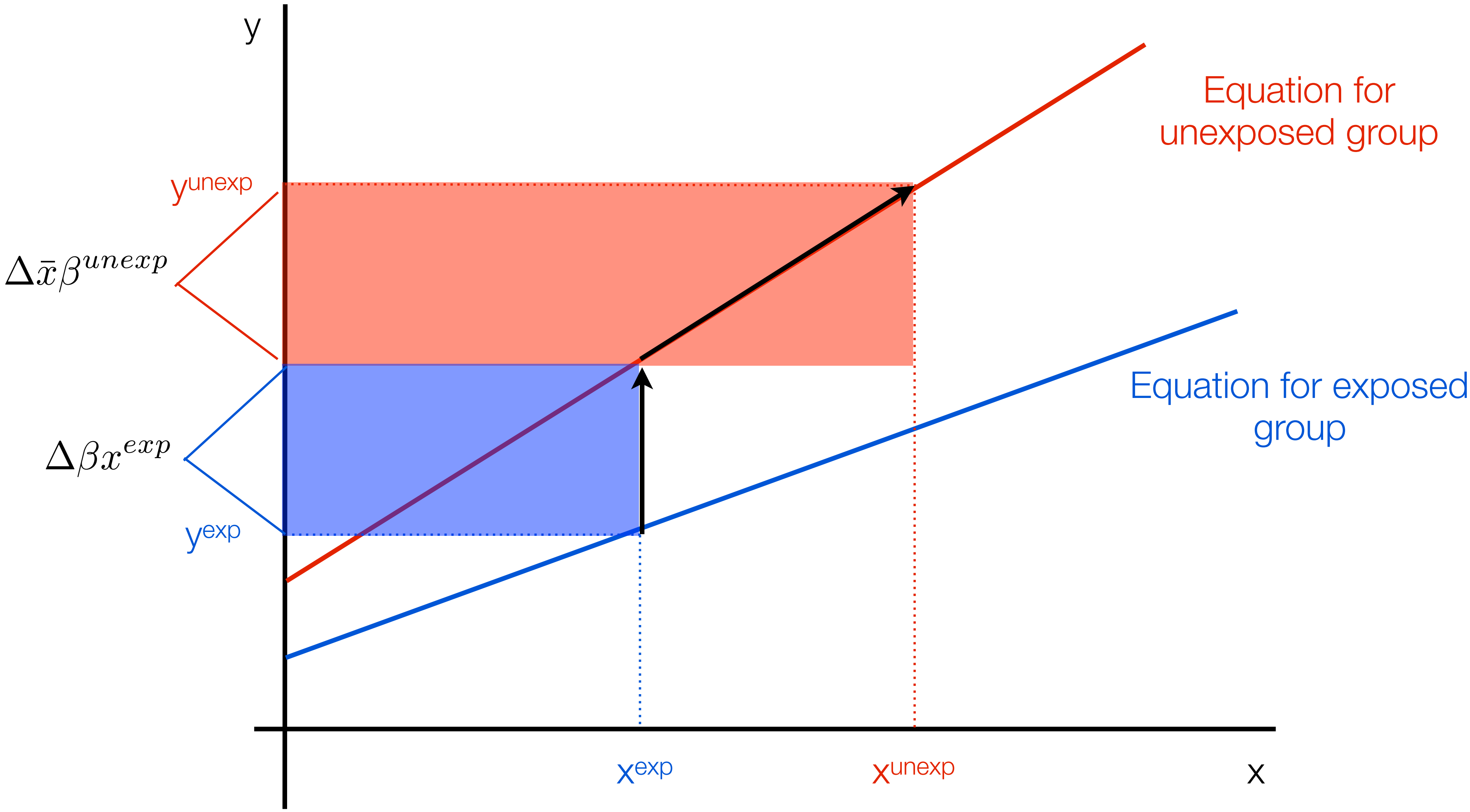
$$y^{exp} - y^{unexp} = \Delta\bar{x}\beta^{unexp} + \Delta\beta x^{exp}$$

$$\text{where } \Delta\bar{x} = \bar{x}^{exp} - \bar{x}^{unexp} \ \ and \ \ \Delta\beta = \beta exp - \beta unexp$$

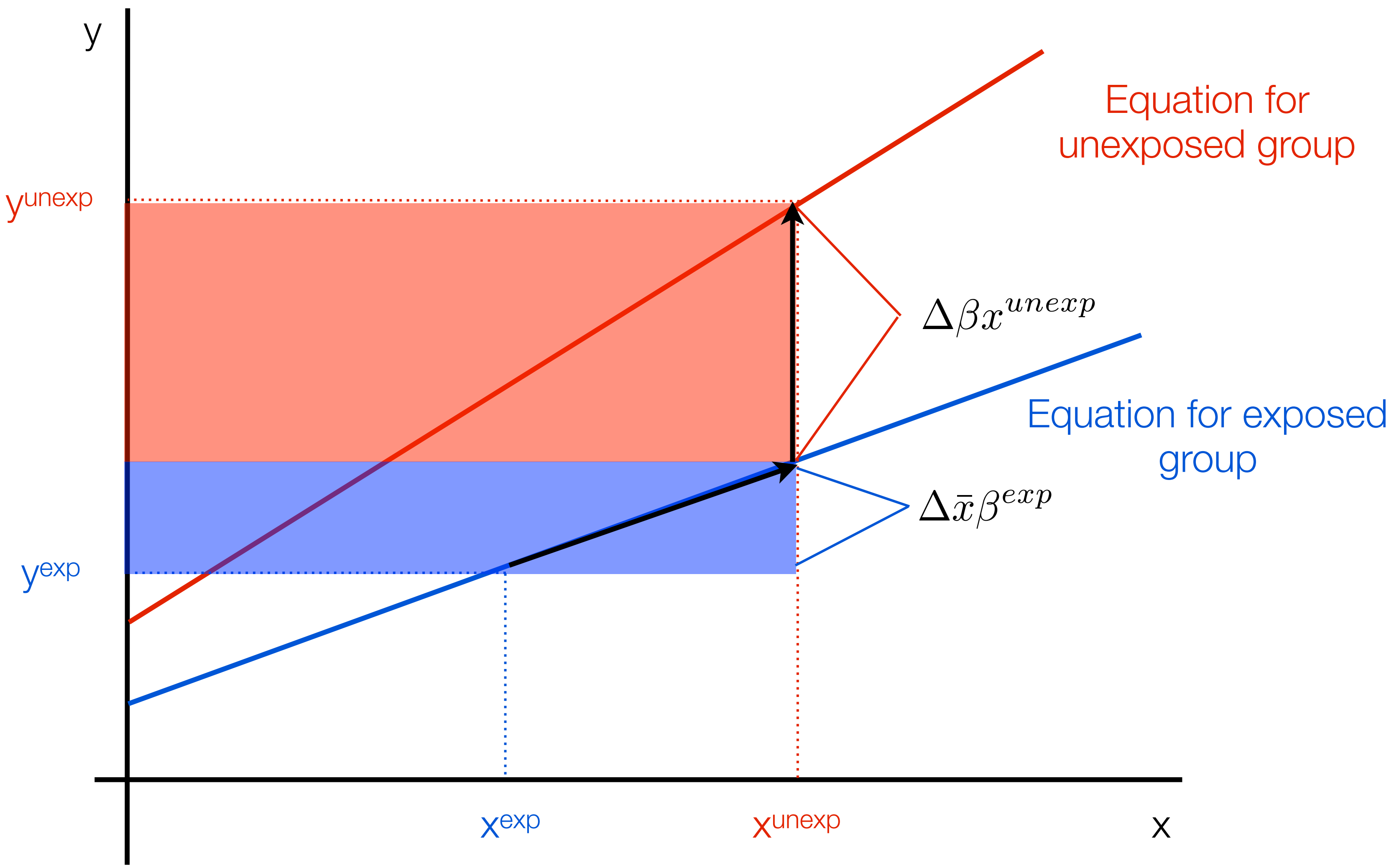
or, equivalently:

$$y^{exp} - y^{unexp} = \Delta\bar{x}\beta^{exp} + \Delta\beta x^{unexp}$$

First method: $y^{exp} - y^{unexp} = \Delta\bar{x}\beta^{unexp} + \Delta\beta x^{exp}$



Second method: $y^{exp} - y^{unexp} = \Delta\bar{x}\beta^{exp} + \Delta\beta x^{unexp}$



- The two methods are equally valid

- In the first,the differences in the x's are weighted by the coefficients of the unexposed group and the differences in the coefficients are weighted by the x's of the exposed group:

$$y^{exp} - y^{unexp} = \Delta \bar{x} \beta^{unexp} + \Delta \beta x^{exp}$$

- whereas, in the second, the differences in the x's are weighted by the coefficients of the exposed group and the differences in the coefficients are weighted by the x's of the unexposed group.

$$y^{exp} - y^{unexp} = \Delta \bar{x} \beta^{exp} + \Delta \beta x^{unexp}$$

- General decomposition formula shows the mean gap as deriving from a difference in endowments (E), a gap in coefficients (C), and a gap arising from the interaction of endowments and coefficients (CE):

$$\begin{aligned} y^{exp} - y^{unexp} &= \Delta\bar{x}\beta^{exp} + \Delta\beta x^{exp} + \Delta\bar{x}\Delta\beta \\ &= E + C + CE \end{aligned}$$

- Method 1 includes interaction with “explained” part:

$$\begin{aligned} y^{exp} - y^{unexp} &= \Delta\bar{x}\beta^{unexp} + \Delta\beta x^{exp} \\ &= (E + CE) + C \end{aligned}$$

- Method 2 includes interaction with “unexplained” part:

$$\begin{aligned} y^{exp} - y^{unexp} &= \Delta\bar{x}\beta^{exp} + \Delta\beta x^{unexp} \\ &= E + (CE + C) \end{aligned}$$

Example: Decomposing Educational Differences in Blood Pressure

Basic question

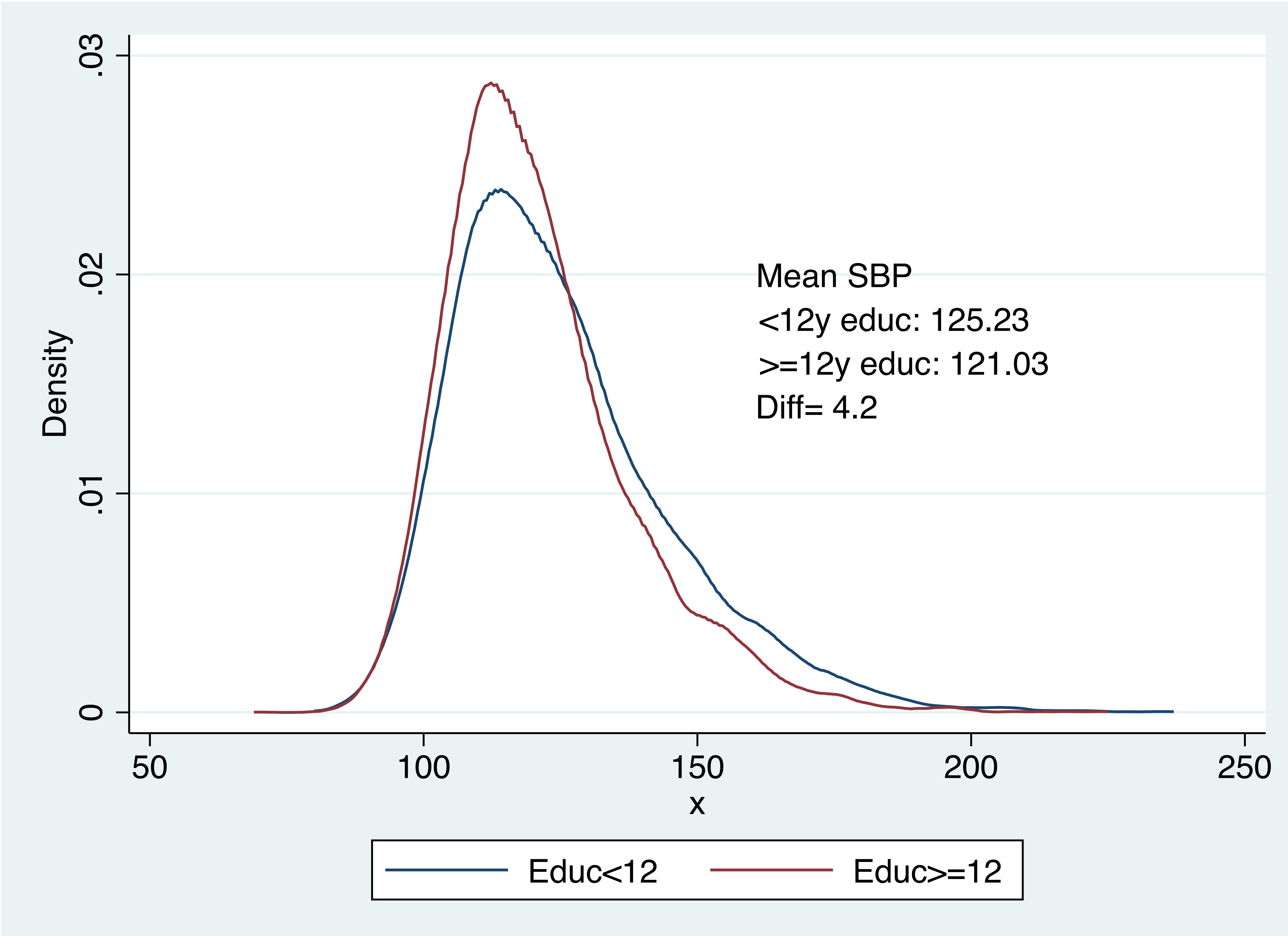
- What is the average difference in blood pressure between those with low vs. high education?

- How much of this difference is due to the fact that determinants of blood pressure (e.g., BMI, smoking, demographics) differ between low and high educated groups?

- Any residual difference is due to educational differences in the associations of risk factors for blood pressure.

Example data

- US NHANES follow up survey (1988-2006), baseline data
- Systolic blood pressure as outcome (mmHg)
- Overall difference by education (0: >=12y educ, 1: <12y educ)
- Potential determinants (the Xs)
  - age (years)
  - age squared
  - race (1 = non-white, 0 = other)
  - marital status (1=married, 0=other)
  - body mass index (kg/m^2)
  - smoking (1=current smoker, 0=other)


Mean SBP
<12y educ: 125.23
>=12y educ: 121.03
Diff= 4.2
Density
0
.01
.02
.03
50
100
150
200
250
x
Educ<12
Educ>=12

Differences in determinants

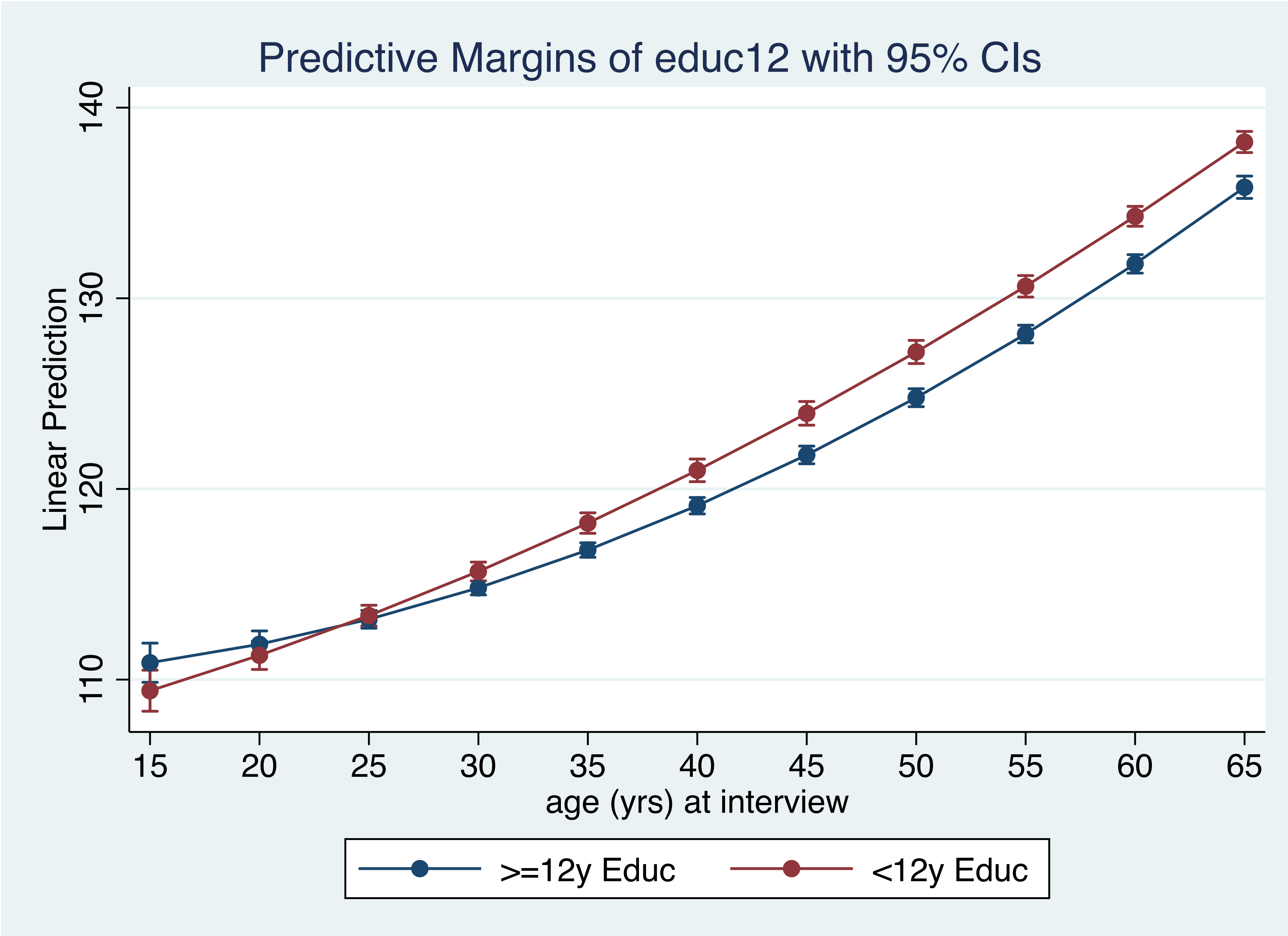
- Lower educated have higher BMI and are more likely to be smokers, as well as being older

	Covariate means			
	<12y Educ		>=12y Educ	
Variable	$\overline{x}$	SD($\overline{x}$)	$\overline{x}$	SD($\overline{x}$)
Age	44.6	18.7	40.9	15.8
Age*Age	2338	1705	1920	1436
Non-white	0.33	0.47	0.36	0.48
Married	0.42	0.49	0.40	0.49
BMI	27.4	5.6	26.9	5.6
Smoker	0.31	0.46	0.25	0.43

Differences in coefficients

- BMI and smoking both have *larger* coefficients for the better educated group.
- Age has a slightly stronger association for the less educated.

	Regression coefficients			
	<12y Educ		>=12y Educ	
Variable	β	SE(β)	β	SE(β)
Age	0.60	0.01	0.53	0.01
Age*Age	0.00	0.00	0.01	0.00
Non-white	2.17	0.44	2.43	0.31
Married	0.92	0.44	0.89	0.32
BMI	0.38	0.04	0.61	0.02
Smoker	0.73	0.44	1.10	0.33
Intercept	110.86	1.11	102.20	0.74


Predictive Margins of educ12 with 95% CIs
Linear Prediction
140
130
120
110
15
20
25
30
35
40
45
50
55
60
65
age (yrs) at interview
>=12y Educ
<12y Educ

SBP (mmHg)	Coefficients used in decomposition: <12y Educ Est.	<12y Educ SE	>=12y Educ Est.	>=12y Educ SE	Pooled Est.	Pooled SE
>=12y Educ	125.23	0.25	125.23	0.25	121.03	0.17
<12y Educ	125.23	0.25	125.23	0.25	125.23	0.25
Difference	-4.20	0.30	-4.20	0.30	-4.20	0.30
Δ due to:						
Covariate Means	**-2.77**	**0.20**	**-2.88**	**0.19**	**-2.85**	**0.19**
Age	-2.14	0.17	-1.89	0.16	-2.00	0.16
Age*Age	-0.46	0.08	-0.69	0.07	-0.59	0.06
Non-white	0.07	0.02	0.07	0.02	0.07	0.02
Married	-0.02	0.01	-0.02	0.01	-0.02	0.01
BMI	-0.18	0.04	-0.29	0.06	-0.25	0.05
Smoker	-0.04	0.03	-0.06	0.02	-0.06	0.02
Coefficients	**-1.29**	**0.25**	**-1.40**	**0.26**	**-1.32**	**0.25**
Age	-0.13	0.03	0.11	0.03	-0.02	0.01
Age*Age	0.79	0.35	0.56	0.25	0.69	0.32
Non-white	0.08	0.18	0.09	0.19	0.08	0.19
Married	-0.01	0.23	-0.01	0.21	-0.01	0.23
BMI	0.06	0.02	-0.05	0.02	0.02	0.01
Smoker	0.11	0.17	0.09	0.14	0.11	0.16
Intercept	-2.20	0.48	-2.20	0.48	-2.20	0.47
Interaction	**-0.11**	**0.11**	**0.11**	**0.11**		

Contribution of covariate differences → Covariate Means

Contribution of coefficient differences → Coefficients

Interaction between coefficients and covariates → Interaction

Coeffi

SBP (mmHg)	<12y Educ	
	Est.	SE
>=12y Educ	125.23	0.25
<12y Educ	125.23	0.25
Difference	-4.20	0.30
Δ due to:		
Covariate Means	**-2.77**	**0.20**
Age	-2.14	0.17
Age*Age	-0.46	0.08
Non-white	0.07	0.02
Married	-0.02	0.01
BMI	-0.18	0.04
Smoker	-0.04	0.03
Coefficients	**-1.29**	**0.25**
Age	-0.13	0.03
Age*Age	0.79	0.35
Non-white	0.08	0.18
Married	-0.01	0.23
BMI	0.06	0.02
Smoker	0.11	0.17
Intercept	-2.20	0.48
Interaction	**-0.11**	**0.11**

Contribution of covariate differences

SBP among the low educated group would be 2.8 mmHg lower if they had the same covariate characteristics as the higher educated.

Most of this difference comes from differences in the distribution of age.

Why positive? This means that the SBP difference would be even larger if the low educated had the same percentage non-white as the higher educated.

SBP (mmHg)	Coeffi <12y Educ Est.	 SE
>=12y Educ	125.23	0.25
<12y Educ	125.23	0.25
Difference	-4.20	0.30
Δ due to:		
Covariate Means	**-2.77**	**0.20**
Age	-2.14	0.17
Age*Age	-0.46	0.08
Non-white	0.07	0.02
Married	-0.02	0.01
BMI	-0.18	0.04
Smoker	-0.04	0.03
Coefficients	**-1.29**	**0.25**
Age	-0.13	0.03
Age*Age	0.79	0.35
Non-white	0.08	0.18
Married	-0.01	0.23
BMI	0.06	0.02
Smoker	0.11	0.17
Intercept	-2.20	0.48
Interaction	**-0.11**	**0.11**

Contribution of coefficient differences

SBP among the low educated group would be 1.3 mmHg lower if they had the same regression coefficients as the higher educated.

Most of this difference is captured by the intercept (i.e., unmeasured factors).

Why positive? This means that the SBP difference would be even larger if smoking had the same effect in low educated as it does in the higher educated.

Similar results if we use the coefficients of the higher educated to weight the covariate differences

	Coefficients used in decomposition:					
	<12y Educ		>=12y Educ		Pooled	
SBP (mmHg)	Est.	SE	Est.	SE	Est.	SE
>=12y Educ	125.23	0.25	125.23	0.25	121.03	0.17
<12y Educ	125.23	0.25	125.23	0.25	125.23	0.25
Difference	-4.20	0.30	-4.20	0.30	-4.20	0.30
Δ due to:						
Covariate Means	**-2.77**	**0.20**	**-2.88**	**0.19**	**-2.85**	**0.19**
Age	-2.14	0.17	-1.89	0.16	-2.00	0.16
Age*Age	-0.46	0.08	-0.69	0.07	-0.59	0.06
Non-white	0.07	0.02	0.07	0.02	0.07	0.02
Married	-0.02	0.01	-0.02	0.01	-0.02	0.01
BMI	-0.18	0.04	-0.29	0.06	-0.25	0.05
Smoker	-0.04	0.03	-0.06	0.02	-0.06	0.02
Coefficients	**-1.29**	**0.25**	**-1.40**	**0.26**	**-1.32**	**0.25**
Age	-0.13	0.03	0.11	0.03	-0.02	0.01
Age*Age	0.79	0.35	0.56	0.25	0.69	0.32
Non-white	0.08	0.18	0.09	0.19	0.08	0.19
Married	-0.01	0.23	-0.01	0.21	-0.01	0.23
BMI	0.06	0.02	-0.05	0.02	0.02	0.01
Smoker	0.11	0.17	0.09	0.14	0.11	0.16
Intercept	-2.20	0.48	-2.20	0.48	-2.20	0.47
Interaction	**0.11**	**0.11**	**0.11**	**0.11**		

Using coefficients from a model pooling both groups together also gives similar results.

No interaction term because only one set of coefficients is used for both group predictions.

	Coefficients used in decomposition:					
	<12y Educ		>=12y Educ		Pooled	
SBP (mmHg)	Est.	SE	Est.	SE	Est.	SE
>=12y Educ	125.23	0.25	125.23	0.25	121.03	0.17
<12y Educ	125.23	0.25	125.23	0.25	125.23	0.25
Difference	-4.20	0.30	-4.20	0.30	-4.20	0.30
Δ due to:						
Covariate Means	**-2.77**	**0.20**	**-2.88**	**0.19**	**-2.85**	**0.19**
Age	-2.14	0.17	-1.89	0.16	-2.00	0.16
Age*Age	-0.46	0.08	-0.69	0.07	-0.59	0.06
Non-white	0.07	0.02	0.07	0.02	0.07	0.02
Married	-0.02	0.01	-0.02	0.01	-0.02	0.01
BMI	-0.18	0.04	-0.29	0.06	-0.25	0.05
Smoker	-0.04	0.03	-0.06	0.02	-0.06	0.02
Coefficients	**-1.29**	**0.25**	**-1.40**	**0.26**	**-1.32**	**0.25**
Age	-0.13	0.03	0.11	0.03	-0.02	0.01
Age*Age	0.79	0.35	0.56	0.25	0.69	0.32
Non-white	0.08	0.18	0.09	0.19	0.08	0.19
Married	-0.01	0.23	-0.01	0.21	-0.01	0.23
BMI	0.06	0.02	-0.05	0.02	0.02	0.01
Smoker	0.11	0.17	0.09	0.14	0.11	0.16
Intercept	-2.20	0.48	-2.20	0.48	-2.20	0.47
Interaction	**0.11**	**0.11**	**0.11**	**0.11**		

Caveat: results depend heavily on quality of specification

```
. oaxaca systolic agec agec2 nonwhite married bmic current male, by(educ12) nodetail

Blinder-Oaxaca decomposition                      Number of obs     =     15,859
                                                    Model           =     linear
Group 1: educ12 = 0                                 N of obs 1      =       9532
Group 2: educ12 = 1                                 N of obs 2      =       6327

------------------------------------------------------------------------------
    systolic |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
overall      |
     group_1 |   121.0268   .1744272   693.85   0.000     120.6849    121.3686
     group_2 |   125.1985   .2500719   500.65   0.000     124.7084    125.6886
  difference |  -4.171762   .3048947   -13.68   0.000    -4.769345    -3.57418
  endowments |  -2.949963   .2080375   -14.18   0.000     -3.35771   -2.542217
coefficients |  -1.023872   .2494773    -4.10   0.000    -1.512839   -.5349059
 interaction |  -.1979264   .1126793    -1.76   0.079    -.4187737    .0229209
------------------------------------------------------------------------------
```

Adding gender increases the “explained” component (i.e., “endowments”) from -2.77 to -2.95, so important consequences for how much of the gap is “unexplained”.

Summary

- Various decomposition techniques exist that may be useful for analyzing social determinants of health
    - Life table decomposition—over time or between groups, or both
    - Regression-based decomposition of Concentration Index
    - Oaxaca decomposition of mean health between groups

- All of these techniques make assumptions that need to be evaluated in the course of analysis

- When used properly, decomposition techniques can help to provide key evidence on *why* health inequalities exist and change over time.