Appendix to

"Dying for a smoke: How much does differential mortality of smokers affect estimated life-course smoking prevalence?"

Rebekka Christopoulou, Jeffrey Han, Ahmed Jaber, Dean R. Lillard

Preventive Medicine 52 (2011) 66–70

The Peto et al. method for caclulating smoking-related mortality

For lung cancer, the Peto et al. method involves approximating national non-smoker death rates by the death rates of a reference population of non-smokers. Lung cancer attributable mortality in a given country is thus calculated as the difference between total national death rates and the respective death rates in the reference population. For other smoking-related diseases, we calculate the proportion of smoking-attributable deaths ($PSAD$) as follows:

$$PSAD = \frac{SIR \cdot c \cdot (RR - 1)}{SIR \cdot c \cdot (RR - 1) + 1} \tag{1}$$

SIR stands for Smoking Impact Factor and is a proxy for the proportion of ever-smokers in the country of interest. We use the standard definition for SIR: $SIR = (C - N)/(S - N)$, where C denotes lung cancer death rates in the country of interest, and N and S denote non-smoker and smoker lung cancer death rates in the reference population, respectively. More recent studies adjust the SIR for lung-cancer risk associated with household energy use, specifically with the burning of coal in stoves and buildings with poor ventilation [1-3]. We do not adjust for this here because we lack the relevant data for the early part of our sample period. This omission may introduce some upward bias in SIR, especially in the early years. In more recent years, domestic coal use in unvented stoves is absent or negligible in the countries we investigate [4]. RR denotes death rates of smokers relative to non-smokers for the disease of interest in the reference population.

Finally, c is a constant factor that accounts for potential confounding and extrapolation bias in RR.

Researchers set the value of c at something less than one to give less weight to RR. Peto *et al* used $c = 0.5$. Other evidence, however, suggests that this value is too conservative [5]. Following Ezzati and Lopez [3], we use $c = 0.7$, but also test the sensitivity of the results to a range of correction factors (from a very conservative value of 0.4 to the maximum of 1.0) and find that results do not substantively change. Variations in c cause some variations in the size of the differential mortality bias, but do not substantially affect the tests for statistical significance. When we set c=0.4, the bias for the second-oldest generation of men in Russia switches from being significant at the 5% level to being significant at 10% level. When we set c=1 the bias for the oldest cohort of UK and US women and the second-oldest cohort of UK and US men switches from being statistically insignificant to being marginally significant at the 10% level. We do not report the sensitivity results here, but they are available upon request.

For each year, sex, and age-group, the product of $PSAD$ with the total number of deaths from the disease of interest gives the smoking-attributable number of deaths from this disease.

Regarding the reference population, we follow the literature and use data from the second phase of the Cancer Prevention Study (CPS-II) carried out by the American Cancer Society. M.J. Thun, MD, MS (Vice President Emeritus, Epidemiology and Surveillance Research, American Cancer Society) generously provided us with the CPS-II 1982-88 data via electronic mail (February 2010). The CPS-II sample includes more than 1 million Americans above the age of 30 starting in 1982. Like most studies, we use the CPS-II because it provides mortality rates of some of the first cohorts of men to have smoked heavily and because it conveniently disaggregates them by sex and age.

Test for statistical significance of differences

To test whether differences between the alternative measures of smoking prevalence rates are statistically significant, we use the Pearson χ^2 test. This standard test of independence for binary variables compares the distributions of scores on the smoking variable implied by each measure of prevalence. To be conservative, and because it better approximates the binomial distribution, we use the χ^2-statistic with the Yates correction for small samples [6]. The Yates correction is applied to 2×2 contingency tables with one or more cells with less than 5 observations. We use it because our data include years when very few people smoke and/or small sub-group samples. Because the correction reduces the value of the χ^2-statistic, the Yates-corrected test is more conservative. We calculate the χ^2-statistic, for each gender, birth-cohort, and year, as follows:

$$\chi^2 = \sum_{i=1}^{N=4} \frac{(\mid O_i - E_i \mid -0.5)^2}{E_i} \sim \chi^2(1)$$

$N = 4$ is the number of distinct events implied by the two measures of smoking prevalence that are under comparison. In other words, N represents the four cells of the relevant 2×2 contingency table (variables indexed by $i = 1, 2$ indentify the number of smokers implied by unadjusted and adjusted prevalence rates, respectively; variables indexed by $i = 3, 4$ identify the number of non-smokers implied by the adjusted and unadjusted prevalence rates, respectively). O denotes observed counts and E denotes expected counts. Since a 2×2 table implies 1 degree of freedom ($=$(no. of rows-1)$\times$(no. of columns-1)), the critical value at the 5% level of significance is always 3.84.

A double-humped pattern of the test statistic may arise when adjusted and unadjusted prevalence rates converge and diverge over certain ages.

References

[1] Liu BQ, Peto R, Chen ZM, et al. Emerging tobacco hazards in China: Retrospective proportional mortality study of one million deaths. *BMJ.* 1998; 317:1411-1422.

[2] Ezzati M, Lopez AD. Measuring the accumulated hazards of smoking: global and regional estimates for 2000. *Tob. Control.* 2003; 12:79-85.

[3] Ezzati M, Lopez AD. Estimates of global mortality attributable to smoking in 2000. *Lancet.* 2003; 362:847-852.

[4] Smith KR, Mehta S, Maeusezahl-feuz M. Indoor air pollution from household use of solid fuels. In: Ezzati M, Lopez AD, Murray CJL, Rodgers A, eds. Comparative Quantification of Health Risks: Global and Regional Burden of Diseases Attributable to Selected Major Risks. Geneva, Switzerland: World Health Organization; 2004; 18:1435-1492.

[5] Thun MJ, Apicella LF, Henley SJ. Smoking vs other risk factors as the cause of smoking-attributable mortality: confounding in the courtroom. *JAMA.* 2000; 284:706-712.

[6] Yates F. Contingency table involving small numbers and the X^2 test. *Supplement to the Journal of the Royal Statistical Society.* 1934; 1(2):217-235.

Table A.1: Mapping of smoking-related diseases across ICD classications for UK and US mortality data

Disease category	ICD-10	ICD-9	ICD-8	ICD-7	ICD-6	ICD-5	ICD-4
Neoplasms							
Lip, oral cavity, pharynx	C000-148	B08	A045	A044	140-148	45a-c,e,f	45a-c,e,f
Esophagus	C150-159	B090	A046	A045	150	46a	46a
Stomach	C160-169	B091	A047	A046	151	46b	46b
Colorectal	C180-189	B093-094	A048-049	A047-048	152-154	46c-e	46c,d
Pancreas	C250-259	B096	157	157	157	46g	46f
Larynx	C320-329	B100	A050	A049	161	47a	47a
Trachea, bronchus, lung	C330, C340-349	B101	A051	A050	162-163	47b-f	47b,c
Cervix, uteri	C530-539	B120	A055	A052	171	48	48
Kidney, other urinary	C640–650	189	189	180	180	52a	51a, 53a
Urinary bladder	C670-679	B126	188	181	181	52b,c	51b, 53b
Acute myeloid leukemia	C920-929	205	205	A058	204	74	72
Cardiovascular diseases							
Hypertension	I100-159	B26	A082	A083-084	440-447	131a,102	131,102
Cerebrovascular	I600-698	B29	A085	A070	330-334	83	82
Atherosclerotic, aortic, arterial	I700-789	B300-302	A086	A085	450-456	96-99	96-99
Ischemic	I200–259	B27	A083	A081	420-422	91c, 92a,d,e, 93b,d,e, 94	91b, 92a-c, 93b-e, 94
Rheumatic, pulmonary, other	I000-099, I260-519	B25, B28	A080-081, A084	A079-080, A082	400-402, 410-416, 430-434	58, 90a,b, 91a,b, 92b,c, 93a,c, 95a-c	56, 90, 91a, 93a, 95a-c
Respiratory diseases							
Other respiratory tuberculosis	A150-169	B021	A006	A001	001-008	13	23
Pneumonia, influenza	J100–189	B321-322	A090-092	A088-091	480-483, 490-493	33, 107-109	11, 107-109
Bronchitis, emphysema, asthma	J400-439	B323	A093	A092-093, A095	241, 500-502, 518, 521	106a-c, 110, 112	106a-d, 110, 112
CAO, other chronic pulmonary	J440-449	B325	A096	A097	511-517, 520, 522-527	111, 113-114	111, 113-114

Notes: Harmonization of ICD4-ICD6 classifications has been applied on US data only. Because of changes in the classification of death causes in time, the harmonization is not always accurate. Category 'acute myeloid leukemia' in ICD 7,6,5,4 includes all kinds of leukemia and aleukemia. Category 'CAO, other chronic pulmonary' in ICD 8,7,6,5,4 also includes other respiratory diseases.

Table A.2: Mapping of smoking-related diseases for Russian mortality data

Disease category	Description of equivalent category in the Russian data	Cause no.
Neoplasms		
Lip, oral cavity, pharynx	Malignant neoplasm of lip, oral cavity and pharynx	45
Esophagus	Malignant neoplasm of esophagus	46
Stomach	Malignant neoplasm of stomach	47
Colorectal	Malignant neoplasm of colon	49
Pancreas	Diseases of pancreas	126
Larynx	Malignant neoplasm of larynx	52
Trachea, bronchus, lung	Malignant neoplasm of trachea, bronchus and lung	53
Cervix, uteri	Malignant neoplasm of cervix uteri	58
Kidney, other urinary	Infections of kidney	130
Urinary bladder		
Acute myeloid leukemia	Leukaemia	65
Cardiovascular diseases		
Hypertension	Hypertensive heart disease	86
	Hypertensive renal disease	87
	Hypertensive heart and renal disease	88
	Other and unspecified hypertensive disease	89
	Acute myocardial infarction with hypertensive disease	90
Cerebrovascular	Cerebrovascular disorders with hypertensive disease	98
	Cerebrovascular disorders without hypertensive disease	99
Atherosclerosis, aortic, arterial	Atherosclerotic cardiosclerosis with hypertensive disease	92
	Atherosclerotic cardiosclerosis without hypertensive disease	93
	Unspecified disorders of pericardium, mitral and aortic valves	96
	Diseases of arteries, arterioles and capillaries	100
Ischemic	Other ischaemic heart diseases with hypertensive disease	94
	Other ischaemic heart diseases without hypertensive disease	95
Rheumatic, pulmonary, other	Acute myocardial infraction without hypertensive disease	91
	Other heart diseases	97
Respiratory diseases		
Other respiratory tuberculosis	Respiratory tuberculosis	9
Pneumonia, influenza	Influenza	104
	Viral pneumonia	105
	Pneumococcal pneumonia	106
	Other acute pneumonias	107
	Congenital pneumonia and pneumonia due to aspiration	153
Bronchitis, emphysema, asthma	Chronic bronchitis and emphysema	108
	Asthma	109
CAO, other chronic pulmonary	Accidental inhalation and ingestion causing obstruction of respiratory tract, foreign body entering other orifice	169

Source: Mesle et al.(1996). Downloaded from: http://www.demoscope.ru/weekly/knigi/shkol/shkol.html

Figure A.1: X^2-test of difference between unadjusted and adjusted smoking prevalence derived from retrospective data by country, gender, birth cohort, and year.

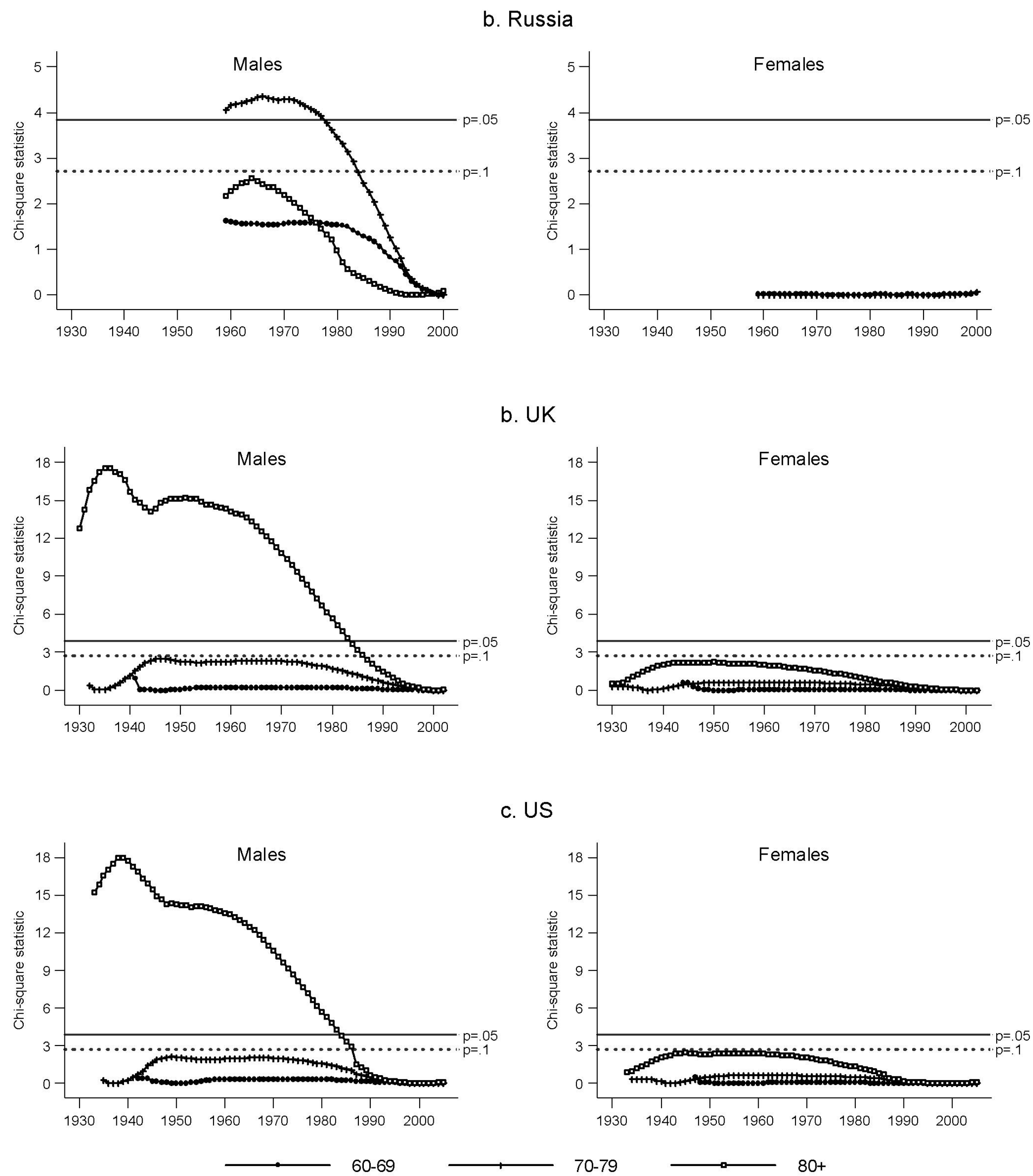

Table A.3: X^2-test of difference between smoking prevalence derived from cross-sectional and retrospective US data

Comparison	A. Cross-sectional vs. unadjusted retrospective data						B. Cross-sectional vs. adjusted retrospective data					
Gender	Males			Females			Males			Females		
Cohort	60-69	70-79	80+	60-69	70-79	80+	60-69	70-79	80+	60-69	70-79	80+
1965	1.6	8.9	0.5	0.2	9.5	1.2	1.0	0.1	16.5	1.2	3.8	1.1
1966	0.5	8.5	3.2	0.6	12.5	0.9	0.1	0.1	9.0	1.0	4.5	1.2
1970	11.2	1.6	0.0	9.9	2.7	0.1	21.6	5.3	22.0	3.8	0.3	2.4
1974	5.3	3.1	0.1	1.8	2.2	0.4	15.7	1.7	16.5	0.1	0.2	1.8
1976	1.6	2.1	0.3	2.7	9.6	1.1	10.6	2.5	12.6	0.3	3.8	0.2
1977	1.9	1.6	0.2	2.7	5.8	2.1	10.4	2.7	10.9	0.6	1.6	0.0
1978	6.8	0.8	0.5	5.0	10.3	0.0	22.5	5.4	8.6	1.8	4.7	2.7
1979	8.9	2.3	0.0	5.4	0.3	0.1	27.5	2.4	10.5	1.8	0.3	1.1
1980	6.2	6.1	0.3	5.5	3.6	1.4	18.1	0.1	7.3	2.9	0.4	0.0
1983	3.0	2.8	0.5	3.8	1.9	0.6	10.7	3.8	6.4	2.7	0.0	0.3
1985	6.2	2.1	0.1	2.5	3.2	0.7	15.2	4.0	4.4	1.8	0.1	0.0
1987	0.3	6.1	0.4	0.3	2.9	2.0	2.1	0.1	2.4	0.4	0.2	0.6
1988	0.8	4.0	1.0	3.8	2.4	2.2	4.2	0.0	0.6	4.1	0.0	0.5
1990	0.2	2.5	5.6	0.5	3.1	1.8	2.5	0.1	0.3	0.0	0.0	0.7
1991	0.3	7.6	6.2	1.8	19.3	5.7	0.0	0.5	1.1	0.5	5.7	4.7
1992	0.2	5.1	6.5	7.3	18.3	6.3	0.0	0.0	1.9	3.3	5.7	4.3
1993	0.1	9.2	3.5	2.5	4.3	7.2	0.0	1.3	0.2	0.6	0.0	4.6
1994	0.0	9.1	5.6	1.3	14.8	5.3	0.1	2.5	1.0	0.0	4.4	3.1
1995	0.3	18.8	9.9	4.6	18.1	8.1	0.0	9.2	3.7	2.4	7.7	4.9
1997	2.8	11.2	1.9	1.3	17.1	6.9	2.0	5.7	0.0	0.1	8.9	3.5
1998	3.4	6.5	0.9	3.6	22.6	6.5	2.1	2.6	0.0	1.2	15.6	2.4
1999	1.5	4.6	0.9	0.4	15.5	11.0	0.3	1.6	0.2	0.1	11.8	7.7
2000	3.2	5.7	0.2	11.5	11.8	6.2	0.6	2.8	0.1	6.7	8.8	2.9
2001	7.0	9.1	1.9	5.3	15.2	4.1	2.8	5.8	0.6	2.4	11.4	1.5
2002	2.1	6.9	2.3	7.5	16.9	4.1	0.4	4.1	1.3	4.6	11.5	1.3
2003	4.5	9.0	0.7	6.9	15.0	4.8	2.3	5.9	0.1	4.1	11.4	1.8
2005	3.6	9.2	1.4	6.1	23.4	3.4	1.1	5.9	0.8	2.4	18.4	1.3

Notes: Critical value at the 5% level is 3.84.