

Supplementary Online Content

Song M, Giovannucci E. Incidence and mortality of carcinoma associated with lifestyle factors among whites in the United States. *JAMA Oncology*. Published online May 19, 2016. doi:10.1001/jamaoncol.2016.0843.

eMethods. Sources of lifestyle data in the US population and derivation of the confidence interval for the population attributable risk (PAR)

eFigure 1. Flowchart of participants' selection in the Nurses' Health Study and Health Professionals Follow-up Study

eTable 1. Number of cancer cases and deaths in the low- and high-risk groups in the Nurses' Health Study (women) and Health Professionals Follow-up Study (men)

eTable 2. Incidence, mortality and the corresponding estimates of population attributable risk of individual cancers in the low- and high-risk groups in the Nurses' Health Study (women) and Health Professionals Follow-up Study (men), as well as in the general US population

This supplementary material has been provided by the authors to give readers additional information about their work.

eMethods

Sources of lifestyle data in the US population

The national data on smoking and physical activity were derived from previous publications using the 2012 National Health Interview Survey, a multipurpose health survey conducted by the Centers for Disease Control and Prevention (CDC).¹ The data on BMI, alcohol consumption, and AHEI score were from the 2009-2010 National Health and Nutrition Examination Survey (NHANES) conducted by CDC.² Details about AHEI derivation in the NHANES have been described elsewhere.³

Derivation of the confidence interval for the population attributable risk (PAR)

The standard error for the national cancer rate was obtained from SEER*Stat software (version 8.1.5) using the method developed by Tiwari *et al.*⁴ We calculated the standard error for the age-adjusted rate in the low-risk group of our cohorts based on Poisson distribution.⁵ PAR was calculated as the difference between the cancer rate of our low-risk group (r_0) and the national SEER rate (r_1) divided by the SEER rate (r_1).

$$\widehat{PAR} = \frac{r_1 - r_0}{r_1} \times 100$$

The variance of PAR can be expressed as

$$\widehat{var}(\widehat{PAR}) = \widehat{var}\left(1 - \frac{r_0}{r_1}\right) = \widehat{var}\left(\frac{r_0}{r_1}\right)$$

Using delta method via log transformation, this variance can be then derived as follows:

$$\widehat{var}(\widehat{PAR}) = \frac{1}{r_1^2} \left[\widehat{var}(r_0) + \left(\frac{r_0}{r_1}\right)^2 \widehat{var}(r_1) \right]$$

In the spirit of transformation suggested by Leung and Kupper,⁶ to improve the asymptotic behavior of the 95% confidence intervals of PAR and to ensure that the confidence intervals remain within the range of -100 to 100, we calculated the confidence intervals using the Fisher's Z transformation,⁷ that is

$$\widehat{var}[Fisherz(\widehat{PAR})] \approx \frac{1}{[(1 + \widehat{PAR})(1 - \widehat{PAR})]^2} \widehat{var}(\widehat{PAR})$$

Then the 95% confidence interval for PAR is estimated as

$$\frac{e^{2[Fisherz(\widehat{PAR}) \pm 1.96\sqrt{\widehat{var}[Fisherz(\widehat{PAR})]}]} - 1}{e^{2[Fisherz(\widehat{PAR}) \pm 1.96\sqrt{\widehat{var}[Fisherz(\widehat{PAR})]}]} + 1},$$

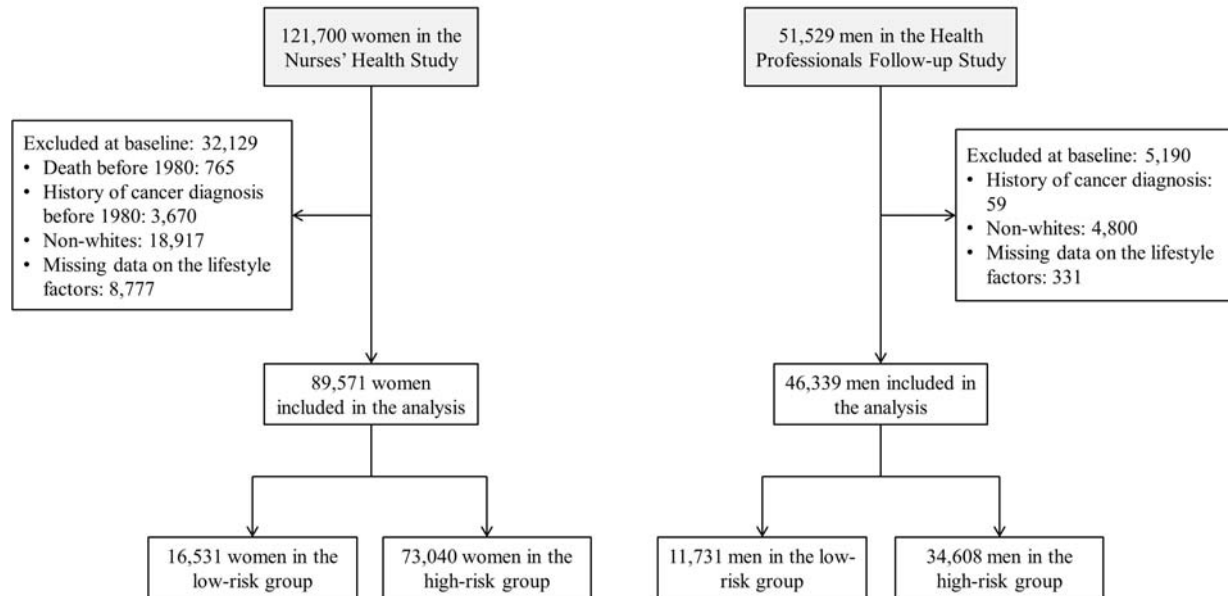
where $Fisherz(\widehat{PAR}) = \log\left[\sqrt{\frac{1+\widehat{PAR}}{1-\widehat{PAR}}}\right]$

References

1. Blackwell DL, Lucas JW, Clarke TC. Summary health statistics for U.S. adults: national health interview survey, 2012. *Vital and health statistics. Series 10, Data from the National Health Survey*. Feb 2014(260):1-161.
2. CDC National Center for Health Statistics. National Health and Nutrition Examination Survey, 2009-2010. http://wwwn.cdc.gov/Nchs/Nhanes/Search/nhanes09_10.aspx. Accessed November 3, 2015.
3. Wang DD, Leung CW, Li Y, et al. Trends in dietary quality among adults in the United States, 1999 through 2010. *JAMA internal medicine*. Oct 2014;174(10):1587-1595.
4. Tiwari RC, Clegg LX, Zou Z. Efficient interval estimation for age-adjusted cancer rates. *Stat Methods Med Res*. Dec 2006;15(6):547-569.

5. Boyle P, Parkin DM. Statistical methods for registries. In: Jensen OM, Parkin DM, MacLennan R, Muir CS, Skeet RG, eds. *Cancer Registration: Principles and Methods*. Lyon: International Agency for Research on Cancer (IARC); 1991:138.
6. Leung HM, Kupper LL. Comparisons of confidence intervals for attributable risk. *Biometrics*. Jun 1981;37(2):293-302.
7. Spiegelman D, Hertzmark E, Wand HC. Point and interval estimates of partial population attributable risks in cohort studies: examples and software. *Cancer Causes Control*. Jun 2007;18(5):571-579.

eFigure 1 Flowchart of participants' selection in the Nurses' Health Study and Health Professionals Follow-up Study



eTable 1 Number of cancer cases and deaths in the low- and high-risk groups in the Nurses' Health Study (women) and Health Professionals Follow-up Study (men)*

Cancer site	Women		Men	
	Low-risk group	High-risk group	Low-risk group	High-risk group
Incident cases				
Lung	104	2,024	92	893
Breast	1,538	5,657	-	-
Colon and rectum	293	1,387	268	666
Endometrium	271	1,171	-	-
Ovary	155	580	-	-
Pancreas	73	328	102	296
Bladder	70	418	174	637
Kidney	48	270	99	259
Oral cavity and pharynx	41	175	31	90
Esophagus	14	111	21	159
Liver	12	61	17	57
Death cases				
Lung	89	1,864	93	848
Breast	226	886	-	-
Prostate	-	-	225	575
Colon and rectum	107	511	146	393
Endometrium	33	202	-	-
Ovary	104	413	-	-
Pancreas	96	439	116	305
Bladder	17	109	46	183
Kidney	20	120	31	119
Oral cavity and pharynx	6	67	14	69
Esophagus	13	84	24	139
Liver	11	69	19	55

eTable 2 Incidence, mortality and the corresponding estimates of population attributable risk of individual cancers in the low- and high-risk groups in the Nurses' Health Study (women) and Health Professionals Follow-up Study (men), as well as in the general US population*

	Rate in the low-risk group (per 100 000)	Rate in the high-risk group (per 100 000)	Rate in the US population (per 100 000) [†]	PAR (% , low-risk vs. high-risk groups) (95% CI) [‡]	PAR (% , low-risk vs. national population) (95% CI) [§]
Women					
Incidence					
Breast	242	251	243	4 (-3, 10)	15 (10, 20)
Lung	17	93	17	82 (77, 86)	85 (81, 88)
Colon and rectum	43	61	43	29 (19, 39)	60 (54, 65)
Endometrium	38	49	39	21 (9, 33)	37 (29, 46)
Ovary	21	27	21	21 (5, 37)	34 (22, 45)
Pancreas	11	15	11	30 (8, 49)	53 (40, 65)
Bladder	12	18	12	36 (15, 54)	49 (33, 62)
Kidney	7	11	7	36 (14, 56)	60 (46, 71)
Oral cavity and pharynx	6	8	6	16 (-17, 46)	58 (41, 71)
Esophagus	2	5	2	62 (36, 80)	61 (35, 78)
Liver	2	3	2	27 (-38, 74)	47 (-4, 79)
Mortality					
Breast	63	40	35	12 (-4, 28)	45 (35, 53)
Lung	84	84	13	84 (80, 88)	84 (80, 88)
Colon and rectum	43	24	17	29 (9, 46)	59 (48, 68)
Endometrium	10	9	5	49 (23, 69)	53 (30, 70)
Ovary	21	18	13	29 (10, 45)	39 (26, 51)
Pancreas	21	19	14	26 (5, 45)	32 (15, 48)
Bladder	5	6	4	38 (-10, 72)	30 (-18, 67)

Kidney	6	5	4	34 (-12, 68)	44 (6, 71)
Oral cavity and pharynx	4	3	1	75 (41, 90)	83 (60, 92)
Esophagus	4	4	2	44 (-7, 76)	41 (-8, 73)
Liver	4	3	2	33 (-25, 75)	50 (5, 79)
Men					
Incidence					
Lung	19	85	20	78 (73, 83)	90 (88, 92)
Colon and rectum	56	70	59	20 (6, 32)	59 (54, 65)
Bladder	36	63	37	44 (32, 54)	62 (55, 69)
Fatal prostate	65	49	39	21 (8, 33)	40 (32, 48)
Kidney	26	27	26	4 (-25, 33)	36 (17, 52)
Oral cavity and pharynx	7	10	7	38 (6, 62)	83 (75, 88)
Pancreas	20	29	22	29 (10, 47)	33 (17, 48)
Esophagus	5	15	5	66 (38, 82)	71 (49, 85)
Liver	3	5	4	32 (-9, 64)	77 (63, 86)
Mortality					
Lung	174	74	17	77 (71, 81)	90 (88, 92)
Prostate	65	49	39	21 (8, 33)	40 (32, 48)
Colon and rectum	56	37	27	26 (10, 41)	52 (43, 59)
Bladder	19	16	8	52 (34, 66)	58 (45, 69)
Kidney	14	11	6	48 (24, 67)	60 (43, 72)
Oral cavity and pharynx	9	6	3	57 (25, 77)	71 (52, 84)
Pancreas	28	28	20	27 (10, 43)	28 (14, 41)

Esophagus	17	12	5	55 (22, 77)	68 (45, 83)
Liver	12	5	3	28 (-13, 60)	72 (57, 83)

Abbreviations: CI, confidence interval; PAR, population attributable risk.

* All rates are standardized based on the age distribution of the U.S. population in 2000.

† US rate was obtained from the Surveillance, Epidemiology, and End Results (SEER) program.

‡ Calculated as the difference in the cancer rates between the low- and high-risk groups divided by the rate of the high-risk group.

§ Calculated as the difference in the cancer rates between the low-risk group and the national rate divided by the national rate.