

## Harvard Report on Cancer Prevention Volume 4: Harvard Cancer Risk Index

G.A. Colditz\*, K.A. Atwood, K. Emmons, R.R. Monson, W.C. Willett, D. Trichopoulos & D.J. Hunter  
for the Risk Index Working Group, Harvard Center for Cancer Prevention  
*Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School and the  
Harvard Center for Cancer Prevention, 677 Huntington Avenue, Boston, MA 02115, USA*

*Key words:* cancer prevention, risk communication, risk estimation.

### Abstract

*Objective:* Prediction of cancer risk is a minor component of current health risk appraisals. Perception of individual cancer risk is poor. A Cancer Risk Index was developed to predict individual cancer risk for cancers accounting for 80% of the cancer burden in the United States.

*Methods:* We used group consensus among researchers at the Harvard Medical School and Harvard School of Public Health to identify risk factors as definite, probable and possible causes of cancer. Risk points were allocated according to the strength of the causal association and summed. Population average risk of cancer and cumulative 10-year risk was obtained from SEER data. Individual ranking relative to the population average was determined. The risk index was evaluated for validity using colon cancer incidence in prospective cohort data.

*Results:* The Harvard Cancer Risk Index provides a broad classification of cancer risk. Validation against cohort data shows good agreement for colon cancer.

*Conclusion:* The Harvard Cancer Risk Index offers a simple estimation of personal risk of cancer. It may help inform users of the major risk factors for cancer and identify changes in lifestyle that will reduce their risk. It offers the potential for tailored health-promotion messages.

### Background

Prediction rules have become widely used in clinical practice to assist medical decision-making when caring for patients with clinical disease, and to counsel patients regarding the likely course of their disease. Use of such prediction rules for counseling aimed at prevention of chronic illnesses is less well developed. The Framingham Heart Study has served as a basis for development of a prediction rule for future risk of coronary heart disease and for stroke [1, 2], and rules have been developed to predict clinical disease based on results of tests such as exercise stress testing [3]. No parallel prediction rule has been developed for overall risk of cancer, in part because of the many types of cancer that may be considered. Gail *et al.* have developed a prediction rule for breast

cancer that is now used in clinical settings to identify women at high risk, and to counsel them [4, 5]. Dupont and Plummer have prepared a computer program to estimate absolute risk of cancer given known relative risk estimates [6].

Mounting evidence indicates that more than 50% of cancer could be prevented if our current knowledge of risk factors were successfully implemented to reduce risk factor prevalence [7]. We therefore undertook the development of a cancer risk index that might aid physicians in counseling patients about their cancer risk, and may serve to educate the public about the relative importance of cancer risk factors. The need for better informing the public about underlying cancer risk is evident from numerous sources. For example, Black and colleagues showed that women greatly overestimate their own breast cancer risk as well as the value of screening mammography [8].

We limited the Harvard Cancer Risk Index to the leading forms of cancer that account for approximately

\* Address correspondence to Graham A. Colditz, Channing Laboratory, 181 Longwood Avenue, Boston, MA 02115, USA; Ph. 617-525-2258; Fax 617-525-2008

80% of cancer incidence (excluding non-melanoma skin cancer). A limited number of cancers are addressed, avoiding undue emphasis being placed on rare cancers that make little contribution to total cancer burden. We aimed for a tool that is both manageable and sufficiently broad to be useful. We chose to focus on cancer incidence rather than mortality because the latter depends also on screening and treatment effectiveness.

While the Harvard Cancer Risk Index Working Group recognizes the necessity of aggressive screening in any comprehensive cancer control program, in most instances screening focuses on an already present cancer that is often preclinical and asymptomatic [9]. With the exception of colon/rectal and cervical cancer screenings that can detect precancerous lesions, and thus help prevent cancer occurrence, the benefits of most cancer screening programs is reduction of disease-specific mortality. Because the focus of the Harvard Cancer Risk Index is the primary prevention of cancer, only screening strategies that contribute to the reduction of cancer incidence are included [10, 11].

In this paper, we summarize the development of the Harvard Cancer Risk Index and report the validation of the approach using data from a prospective cohort for colon cancer as an example. We expect that this tool would likely be used in a clinical setting; it can be self-administered and with computerization may include prevention messages appropriate to the risk profile provided by the user.

## Methods

The Harvard Cancer Risk Index Working Group of the Harvard Center for Cancer Prevention was composed of epidemiologists, clinical oncologists, and other Harvard faculty with quantitative expertise focused on cancer and risk assessment (see Appendix). The group met on a monthly basis for over 18 months to formulate this approach to the Harvard Cancer Risk Index described in this paper.

We first identified the cancers accounting for 80% of incidence in the United States based on the Surveillance Epidemiology and End Results (SEER) published rates; this resulted in 10 cancers for men and 13 for women (Table 1). We then adopted a group consensus process to identify the genetic, environmental, nutritional, and lifestyle factors, as well as major illnesses that are established or likely causes of these cancers. We excluded from consideration screening that is not aimed at reducing cancer incidence (*e.g.* mammography which leads to early diagnosis but not prevention of incident disease). In a manner similar to the International

Table 1. Estimated new cancer cases and deaths for 2000

Cancer	Men		Women	
	Cases	Deaths	Cases	Deaths
Prostate	180,400	37,000		
Breast			182,800	40,800
Lung	89,500	89,300	74,600	67,600
Colon	43,400	23,100	50,400	24,600
Bladder	38,300	8100	14,900	4100
Endometrial			36,100	6500
Non-Hodgkin's lymphoma	31,700	13,700	23,200	12,400
Ovarian			23,100	14,000
Skin melanomas	27,300	4800	20,400	2900
Kidney	18,800	7300	12,400	4600
Cervical			12,800	4600
Leukemia	16,900	12,100	13,900	9600
Stomach	13,400	7600	8100	5400
Pancreatic	13,700	13,700	14,600	14,500

Source: Greenlee RT, Murray T, Bolden S, Wingo PA (2000) Cancer Statistics, 2000. *CA Cancer J Clin Cancer* 50: 7–33.

Agency for Research on Cancer (IARC) criteria [12], the causes were classified as definite, probable, and possible. This allowed us to separate the strength of evidence from the magnitude of the association between a risk factor and a specific cancer. We use only the definite and probable causes in our subsequent risk index. For each exposure, a relative risk was identified as the most likely descriptor of the association in humans. To avoid protracted debate regarding the precise magnitude of the association we used five categories of relative risk (none RR = 0.9 to 1.19; weak RR = 1.2 to <1.5; moderate RR = 1.5 to <3.0; strong RR = 3.0 to <7.0; very strong RR = 7 or more). We chose not to use meta-analysis as a tool for the estimation of the relative risk because up-to-date estimates are available for only a few of the causal associations under consideration and meta-analysis would convey a level of precision greater than can be justified.

Following the approach used for cardiovascular disease, where logistic risk functions are translated into an additive scale, we next translated the relative risk associated with an exposure into a number of cancer risk points (see Table 2). The population average number of points was estimated based on the prevalence of risk factors in the US. The individual score can then be compared to the average for the population, thus placing the risk of cancer in context. These population averages for each cancer correspond to the average cancer risk as reflected in the SEER rates, which approximate the national incidence of cancer in the US. We next developed a seven-level scale to rank cancer risk in relation to this US population average. These levels range from very high risk, to about average

Table 2. Conversion of relative risk to cancer risk points for the Harvard Cancer Risk Index

Relative risk (x)	Association	Symbol <sup>a</sup>	Risk points
$0.9 \cdot x < 1.1$	Not discernible		0
$0.7 \cdot x < 0.9, 1.1 \cdot x < 1.5$	Weak	- or +	5
$0.4 \cdot x < 0.7, 1.5 \cdot x < 3.0$	Moderate	-- or + +	10
$0.2 \cdot x < 0.4, 3.0 \cdot x < 7.0$	Strong	--- or + + +	25
$< 0.2, 7.0 +$	Very strong	---- or + + + +	50

<sup>a</sup> Symbol as used to summarize association in Table 4.

Table 3. Levels of cancer risk defined by the ratio of risk points to population average risk. This gives the factor used to multiply the SEER-derived 10-year estimated risk of cancer to obtain a numerical value for the likelihood of cancer diagnosis during the next 10 years

Cancer risk score divided by population average score	Level of risk	SEER multiplier
< 0	Very much below average risk	0.2
0, or < 0.5	Much below average risk	0.4
0.5 < 0.9	Below average risk	0.7
0.9 < 1.1	About average risk	1.0
1.1 < 2.0	Above average risk	1.5
2.0 < 5.0	Much above average risk	3.0
5.0 or more times the average score	Very much above average risk	5.0

risk and very low risk (see Table 3). Through simple addition and subtraction, each individual calculates his or her own total number of risk points for each cancer, which can then be compared against US population averages. Division by the population average gives a relative score that is translated into a five-category comparative ranking from very low risk to average risk to very high risk.

In addition to this relative ranking, the absolute risk over a specified time can also be derived from the US SEER data. The National Cancer Institute SEER program publishes 10-, 20- and 30-year risk of each cancer conditional on current age. We chose to use the 10-year risk for a time frame that was comprehensible to the user. As summarized in Figure 1, the steps are:

1. We estimate risk points for each risk factor based on a simple, usually dichotomous, response.
2. Sum these for the individual.
3. Divide these by the population average.
4. Multiply the level of risk by the SEER rate to estimate risk of cancer diagnosis over the next 10 years.

To assess the validity of the risk index for colon cancer in men, we conducted a prospective analysis of lifestyle factors and colon cancer incidence using the cohort data from the Health Professionals Follow-up Study. We chose colon cancer as it has many potentially modifiable risk factors. After classifying the exposures in a manner comparable to the definitions in the risk index, we estimated the relative rates of colon cancer occurrence

for risk factors that correspond to the categories in the cancer risk index.

**Results**

The cancers accounting for 80% of cancer incidence (excluding non-melanoma skin cancer) are listed in Table 1. After review of the causes of cancer, we omitted non-Hodgkin's lymphoma and stomach cancer due to limited risk factors that are modifiable. The established or likely causes and the level of relative risk associated with each factor are included in Table 4. For simplicity of administration, and ease of self-completion, continuous or ordered variables have usually been dichotom-

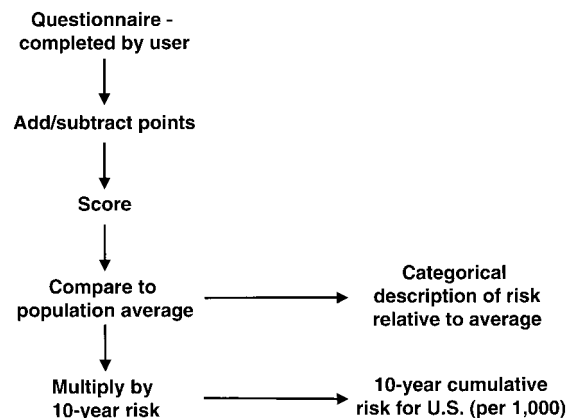


Fig. 1. Flow of risk estimation process.

Table 4. Exposures categorized by strength of evidence and strength of association for specific answers

Strength of evidence	<i>Definite:</i> An association has been established between the exposure and outcome, in which chance, bias and confounding can be ruled out with reasonable confidence			<i>Probable:</i> An association has been observed between the exposure and the outcome but chance, bias and confounding cannot be ruled out with reasonable confidence			<i>Possible:</i> The available studies are of insufficient quality, consistency or statistical power to permit a conclusion of probable or definite association between the exposure and outcome		
	Strength of association			Strength of association			Strength of association		
Prostate	Family history (first-degree relative)	1.8	++	Animal fat <sup>a</sup>	1.8	++	Physical activity (≥3 h/week)	0.8	-
	African-American vs. white	1.8	+++	Height (> 5'10")	1.3	+			
	Asian vs. white	0.4	--	Tomatoes <sup>a</sup>	0.7	-			
Breast				Vasectomy (yes vs. no)	1.5	++			
	Family history (mother and sister) <sup>b</sup>	3.0	+++	Alcohol (> 1 drink/day vs. none)	1.4	+	Monounsaturated fat <sup>a</sup>	0.7	-
	Family history (first-degree relative) <sup>c</sup>	1.8	++	Estrogen replacement (< 5 years vs. no)	1.1	+	Physical activity (≥3 h/week)	0.8	-
	Height (> 5'7")	1.3	+	Breast feeding (> 1 Year vs. none)	0.8	-	Saturated fat	1.2	+
	Age of first period (≥15 vs. ≤11)	0.8	-	Obesity (postmenopausal) > 27 BMI <sup>d</sup> vs. (< 21 BMI)	1.3	+			
	Age first birth (≥35 vs. ≤20)	1.5	++						
	No. of births (0 or 1 child)	1.1	+	Obesity (premenopausal) > 27 BMI <sup>d</sup> vs. (< 21 BMI)	0.8	-			
	Age at menopause (5 year increment)	1.2	+						
	OC use (current use vs. none)	1.4	+	Benign breast disease <sup>c</sup> (MD diagnosed)	1.5	++			
	Estrogen replacement (≥5 years vs. no)	1.7	++	Vegetables <sup>a</sup>	0.8	-			
	Jewish heritage (yes vs. no)	1.2	+						
	Ionizing radiation (yes vs. no)	2.0	++						
	Lung	Smoke (≥25 cigarettes/day vs. none)	10	+++++	Occupational exposure <sup>f</sup>				
		Smoke (15–25 cigarettes/day vs. none)	5.0	++	Air pollution (living in large city vs. no)	1.2	+		
		Smoke (< 15 cigarettes/day vs. none)	2.0	++	Fruits <sup>a</sup>	0.7	-		
Cigar smoking (1 a day for last year)		1.4	+						
Family history (first-degree relative)		1.5	++						
Occupational exposure <sup>f</sup>									
Passive smoking (among non-smokers: yes vs. no)		1.3	+						
Vegetables <sup>a</sup>		0.7	-						

Table 4. (Continued)

Strength of evidence	<i>Definite:</i> An association has been established between the exposure and outcome, in which chance, bias and confounding can be ruled out with reasonable confidence			<i>Probable:</i> An association has been observed between the exposure and the outcome but chance, bias and confounding cannot be ruled out with reasonable confidence			<i>Possible:</i> The available studies are of insufficient quality, consistency or statistical power to permit a conclusion of probable or definite association between the exposure and outcome		
	Strength of association			Strength of association			Strength of association		
Colon	Family history (first-degree relative)	1.8	++	Vegetables <sup>a</sup>	0.7	-	Fruits <sup>a</sup>	0.8	-
	Obesity (> 27 BMI vs. < 21) <sup>d</sup>	1.5	++	Alcohol (> 1 drink/day vs. 0)	1.4	+	Fiber <sup>a</sup>	0.7	-
	Screening (FOBT or sigmoidoscopy vs. none) <sup>e</sup>	0.5	--	Height (6" increment)	1.3	+	Saturated fat <sup>a</sup>	1.4	+
	Aspirin (15 years of regular use)	0.7	-	Physical activity (≥3 h/week vs. none)	0.6	--	Smoke (≥25 cigarettes/day vs. none)	1.5	++
	Inflammatory bowel disease (diagnosed for more than 10 years) <sup>i</sup>	1.5	++	Estrogen replacement (≥5 yrs vs. 0)	0.8	-			
	Folate	0.5	--	OC use (≥5 yrs vs. none)	0.7	-			
				Red meat <sup>a</sup>	1.5	++			
Bladder	Occupational exposure <sup>h</sup>			Family History (first-degree relative)	1.5	++	Vegetables <sup>a</sup>	0.7	-
	Smoke (≥25 cigarettes/day vs. none)	3.0		Occupational exposure <sup>f</sup>			Fruit <sup>a</sup>	0.8	-
	Smoke (15–24 cigarettes/day vs. none)	2.0					Water chlorination (high vs. low/no)	2.0	++
	Smoke (≤14 vs. none)	1.4							
Endometrial	Obesity (> 27 BMI vs. < 21) <sup>d</sup>	2.0	++	Smoke (≥25 cigarettes/day vs. 0)	0.7	-	Vegetables <sup>a</sup>	0.7	-
	Nulliparous (0 vs. ≥1 child)	1.2	+	Family history (first-degree relative)	1.5	++	Fruits <sup>a</sup>	0.8	-
	Age menopause (5 year increment)	1.2	+	Diabetes type II (yes vs. no)	1.5	++	Saturated Fat <sup>a</sup>	1.4	+
	OC use (≥5 years vs. none)	0.5	--				Age of first period (≥15 vs. ≤11)	0.9	-
	Estrogen replacement (≥10 years vs. none)	4.0	+++						
Non-Hodgkin's Lymphoma	HIV infection	> 10	++++				Smoking (high)	1.4	+
	Immunosuppressive drugs	10.0	++++				UV exposure	1.5	++
	Family history (first-degree relative)	1.5					Blood transfusion	1.5	++
Ovarian	Family history (first-degree relative)	1.5	++	Breast feeding (> 1 year vs. none)	0.9	-	Vegetables <sup>a</sup>	0.8	-
	No. of children (0 or 1)	1.3	+				Fruits <sup>a</sup>	0.7	-
	OC use (≥5 years vs. none)	0.7	-				Ovulation inducing drugs (yes vs. no)	0.3	++
	Tubal ligation (yes vs. no)	0.6	--						
	Hysterectomy (yes vs. no)	0.8	-						

Table 4. (Continued)

Strength of evidence	<i>Definite:</i> An association has been established between the exposure and outcome, in which chance, bias and confounding can be ruled out with reasonable confidence			<i>Probable:</i> An association has been observed between the exposure and the outcome but chance, bias and confounding cannot be ruled out with reasonable confidence			<i>Possible:</i> The available studies are of insufficient quality, consistency or statistical power to permit a conclusion of probable or definite association between the exposure and outcome		
	Strength of association			Strength of association			Strength of association		
Kidney	Family history (first-degree relative)	1.5	++	Smoke ( $\geq 25$ cigarettes/day vs. none)	2.0	++	Vegetables <sup>a</sup>	0.8	-
				Smoke (15–24 cigarettes/day vs. none)	1.4	+	Fruits <sup>a</sup>	0.7	-
				Obesity ( $> 27$ BMI vs. $< 21$ ) <sup>d</sup>	2.0	++	Blood transfusion	1.5	++
Leukemia	Therapeutic radiation	2.0	++	Smoke ( $\geq 25$ cigarettes/day vs. none)	1.3	+			
	Occupational exposure <sup>f</sup>	> 10		Occupational exposure <sup>f</sup>	5.0				
Cervical	Age first intercourse ( $< 16$ vs. $\geq 22$ )	1.5	+++	Smoking ( $\geq 25$ cigarettes/day vs. none)	1.3	+	Vegetables <sup>a</sup>	0.8	-
	Multiparous ( $\geq 2$ children vs. $< 2$ )	1.5	+	History of any STD <sup>j</sup>	2.0	++	Fruits <sup>a</sup>	0.7	-
	Multiple sex partners ( $\geq 6$ vs. $\leq 1$ )	1.5	++				Partner circumcision (yes vs. no)	0.5	--
	Barrier methods (as a dominant method of birth control)	0.5	--				OC use ( $> 5$ years vs. none)	2.0	++
	Screening (no Pap smear in the last 3 years vs. yes)	2.0	++				DES (maternal prenatal exposure)	2.0	++
	Low SES ( $< \$10,000$ vs. $> \$10,000$ )	3.8	+++						
Pancreatic	Smoke ( $\geq 25$ cigarettes/day vs. none)	2.5	++	Vegetables <sup>a</sup>	0.6	--	Fruits <sup>a</sup>	0.8	-
	Smoke (15–24 cigarettes/day vs. none)	1.4	+	Diabetes type II (yes vs. no)	2.0	++	Carbohydrates <sup>a</sup>	1.3	+
	Family history (first degree relative)	1.5	++	Chronic pancreatitis (yes vs. no)	4.0	+++	Fiber <sup>a</sup>	0.8	-
Skin melanoma	UV exposure (repeated sun burns)	3.7	++++	Immunosuppressive drugs	2.0	++	Alcohol	1.7	++
	Hair color (red/blonde vs. black)	2.0	+						
	Eye color (blue/green vs. all others)	1.5	++						
	Skin color (light vs. dark/olive)	1.5	++						
	Family history (first degree relative)	2.0	++						

Table 4. (Continued)

Strength of evidence	<i>Definite:</i> An association has been established between the exposure and outcome, in which chance, bias and confounding can be ruled out with reasonable confidence			<i>Probable:</i> An association has been observed between the exposure and the outcome but chance, bias and confounding cannot be ruled out with reasonable confidence			<i>Possible:</i> The available studies are of insufficient quality, consistency or statistical power to permit a conclusion of probable or definite association between the exposure and outcome		
	Strength of association			Strength of association			Strength of association		
Stomach	Low SES (< \$10,000 vs. > \$10,000)	1.8	+++	Fruits <sup>a</sup>	0.5	--	Carbohydrates <sup>a</sup>	1.3	+
	Blood group A	1.2	++	Salt <sup>a</sup>	1.4	+			
	Family history (first-degree relative)	1.5	++						

<sup>a</sup> Upper quartile (top 25%) vs. lower quartile (lower 25%).  
<sup>b</sup> Two first-degree relatives who have a history of breast cancer before age 65 vs. none.  
<sup>c</sup> First degree relative who has a history of breast cancer before age 65 vs. none.  
<sup>d</sup> BMI (body mass index).  
<sup>e</sup> Clinically recognized chronic cystic, fibrocystic or other benign breast vs. none.  
<sup>f</sup> See text.  
<sup>g</sup> Screening at > 50 years of age vs. no screening.  
<sup>h</sup> Lower quintile (lower 25%) vs. upper quintile (upper 25%).  
<sup>i</sup> Physician diagnosed Crohn's disease, ulcerative colitis or pancolitis.  
<sup>j</sup> With unknown HPV status.

ized, though for diet we consider extreme quartiles of the distribution. Occupational exposures are not listed in Table 4. They are, however, incorporated into the index for those who have worked in the production of asbestos, radon, arsenic smelting, production of cadmium, chromium, or beryllium, coal gasification or aluminum production, coke production, iron and steel founding, or exposure to silica without adequate protection. Risk is dependent on two categories of duration of work (more than 20 years or 5–20 years).

We next developed the cancer risk point scale to translate relative risks into risk points. The scale is presented in Table 2. The highest allowed risk point was 50 for a relative risk of 7 or more. This is the level of risk that a smoker of 25 or more cigarettes per day would have for lung cancer, or a worker exposed to asbestos without protection for 20 or more years would have for lung cancer. Strong associations in the range of 3–7 were assigned 25 points. This is the level of risk that a smoker of 15–24 cigarettes per day would have for lung cancer.

To estimate the population average risk points for each cancer, the current US prevalence of each definite or probable cause was multiplied by the risk points for that cause. To illustrate this approach, the data for colon cancer are summarized in Table 5; average population risk was calculated to be 16 points. To compare risk of a certain cancer for an individual against the SEER derived population average, we divided the risk points for the particular individual by the population

average. The cut-off points for levels of risk compared to the average are indicated in Table 3.

*Scenario 1.* The first scenario we consider is that of a 50-year-old woman who has not had a sigmoidoscopy in the past 10 years, has no family history of colorectal cancer, is not obese, drinks less than one alcoholic beverage per day, eats fewer than three servings of vegetables per day, is less than 5 feet 7 inches tall, is not physically active, and eats red meat less often than daily. From Table 4, we see that for such a person we obtain a score of 10. We divide this by the population average score (10/16) and obtain a value of 0.6. This value is between 0.5 and 0.9, so we classify the risk as *below average* (Table 3).

To convert this level of risk to a numeric estimate of colon cancer risk, we look up the multiplier in Table 3 – it is 0.7. That is, we estimate that the risk is 0.7 times the population average risk of colon cancer. Then we multiply the SEER (average) risk of colon cancer by 0.7. For a 50-year-old woman, the 10-year risk of colon cancer is 0.6% in the next 10 years. Multiplying this by 0.7 gives a risk of 0.4% or approximately four cases in 1000 women in 10 years.

*Scenario 2.* The woman for this scenario has a positive family history of colon cancer, has not had a screening sigmoidoscopy in the past 10 years, and does not score points for other risk factors in listed in Table 2. The score for this person is 10 (for family history) plus 10

Table 5. Estimating population average cancer risk points for colon cancer among US women<sup>a</sup>

Risk factor	RR	Description	Risk points	US prevalence	Population average points
Family history	1.8	Brother, sister or parent had colon cancer	10	0.05	0.5
Obesity	1.5	27 kg/m <sup>2</sup> or more	10	0.40	4.0
Screening	2.0	No flexible sigmoidoscopy or FOBT	10	0.76	7.6
Alcohol	1.4	More than 7 servings per week	5	0.02	0.1
Vegetables	0.7	3 or more servings per day	-5	0.25	-1.25
Height	1.3	5 ft 7 in or taller	5	0.10	0.5
Physical activity	0.6	3 or more hours total leisure-time physical activity per week	-10	0.19	-1.9
Red meat (upper quintile)	1.5	Eating 7 or more servings a week	10	0.25	2.5
Use of birth control pills	0.7	5 or more years of use	-5	0.20	-1.0
Use of postmenopausal hormones	0.8	5 or more years of use	-5	0.07	-0.35
Aspirin use	0.7	Use daily for 15 years or more	-5	0.13	-0.65
Multivitamin (folate)	0.5	Low folate intake reflected in lack of regular multivitamin use	10	0.60	6.0
Inflammatory bowel disease	1.5	10 or more years	10	0.001	0.01

<sup>a</sup> For men US prevalences vary and use of oral contraceptives and postmenopausal hormones omitted.

(for no sigmoidoscopy) = 20 points. If we divide this by the population average score of 16 we get 1.2, and turning to Table 3 we see that the woman is at *above average risk* for colon cancer. To convert this risk from the descriptive statement to a numeric estimate we see from Table 5 that we must multiply the SEER or average risk of colon cancer by 1.5. For a 50-year-old woman the risk of colon cancer is 0.6% in the next 10 years. Multiplying this by 1.5 gives a risk of 0.9%, or approximately one case in 100 women in 10 years.

*Scenario 3.* What would happen to the level of risk if the person in Scenario 2 had a sigmoidoscopy? We now consider a third scenario that is the same as number 2 above, except that a sigmoidoscopy has been performed. Now the score is 0. Table 3 tells us that the risk is *much below average*. To obtain a numeric estimate of risk for such a person we would multiply the SEER risk by 0.4. Given the epidemiologic evidence that sigmoidoscopy reduces risk of colon cancer incidence and mortality by approximately 30% over 10 years after the test, this estimate of risk seems appropriate.

*Scenario 4.* In this scenario, we consider the further reduction in risk that such a person may achieve by exercising, at the level of 3 or more hours of physical activity per week. For such a person the score is minus 10. Table 3 tells us that the risk is now *very much below average* because the score is less than zero. The numeric estimate for such a scenario would be 0.2 times the SEER risk.

As a final verification of the colon cancer scale, we can see which exposures would be required to obtain a negative score (corresponding to the lowest level of risk).

To obtain a negative score one could have no risk factors that add points, as shown in Table 5, and eat three servings or more of vegetables per day. Alternatively, one can be physically active by exercising 3 or more hours per week (minus 10 points) and be at very low risk even if he or she has one of the risk factors that conveys five points (for instance, be more than 5 feet 10 inches tall, or drink more than one alcoholic beverage per day).

#### Validation

To assess the validity of the Harvard Cancer Risk Index, we used prospective data from the Nurses' Health Study and Health Professionals Follow-up Study after classifying the lifestyle factors according to the definitions used for colon cancer in Table 5. The incidence rate ratio for combinations of lifestyle factors estimated from the multivariate logistic function shows good agreement with the values developed by the group process. These are summarized in Table 6.

#### Risk of any cancer

Because numerous causes of cancer operate on more than one cancer, and individuals may be concerned about a diagnosis of cancer in general, we developed a scale for "total cancer". This scale is based on the contribution of each of the cancers that we consider to be the grand total number of cases diagnosed for these cancers. For total cancer risk, the "risk points" for an exposure are the weighted average number of points overall where the weight is the proportion of cancer contributed by each site. This weighting assumes that the risk factors for the cancers in the risk index apply



Table 6. Validation of the risk index against prospective data on colon cancer among men

Colon cancer risk profile	Risk index points	Observed RR	Scale value
(a) Population average	18.1	1.0	1.0
(b) No risk factors, no sigmoidoscopy	10	0.75	0.7
(c) No risk factors, recent sigmoidoscopy	0	0.55	0.4
(d) No risk factors, recent sigmoidoscopy and exercise	Negative	0.4	0.2
(e) Obese, no screen, red meat, alcohol	35	2.3	3.0
(f) Obese, no screen, red meat, family history and low folate intake	50	4.0	3.0

proportionately to the 20% of total cancer that we did not specifically consider. This scale gives an overall summary of total cancer risk and may be useful in drawing attention to the overall importance of lifestyle factors such as cigarette smoking and diet.

## Discussion

We have developed a cancer risk index drawing on expert opinion and evaluated it with prospective cohort data. This approach meets the guidelines recommended by Wasson *et al.* and modified by Laupacis *et al.* [13, 14] for the development of prediction rules. While prediction rules have been shown to alter clinical practice [15], there is little such evidence with respect to health behavior. However, we note that counseling by health-care providers to modify lifestyle—including cigarette smoking, diet, and physical activity – is recommended by the US Preventive Services Task Force [9]. Strong evidence indicates that counseling by physicians is effective in smoking cessation. This cancer risk index may help focus counseling by health-care providers on modifiable factors. Further work is needed to evaluate the utility of the index as an aid to tailored counseling for lifestyle change and cancer risk reduction.

One concern with an index such as this is the potential for inappropriate precision to be conveyed to the user. False reassurance should clearly be avoided. We have tied the risk estimates to the US national data and present them for 10-year age groups. Computerized administration would allow prediction based on actual current age. Whether 10-year risk of cancer is the best metric remains uncertain [16, 17]. We have chosen to present the risk of diagnosis within the next 10 years because risk of major illness in the distant future is not a meaningful probability for most people. Further, lifetime risk is not well understood [18]. Given the SEER database, however, alternative time frames could be used. Moreover, as the time frame lengthens the absolute risk increases. Further, long-term predictions of future cancer incidence would require correct mod-

eling of cohort effects and time trends in cancer incidence. For example colon cancer incidence has been decreasing for the past 20 years, particularly among women, making long-term prediction problematic for this cancer which is the second leading cause of cancer mortality [19]. Also, predicting the effect of changing an exposure introduces additional uncertainties because the rapidity with which the change in risk occurs will vary among exposures and cancers.

The Harvard Cancer Risk Index offers both a verbal description of cancer risk and a quantitative estimate of being diagnosed with cancer in the next 10 years. The quantitative estimate is tied to the SEER rates, reflecting the contemporary experience of cancer rates in the United States. The likelihood of cancer diagnosis is presented as a risk in 1000. This can be interpreted as the number of people in a group of 1000, of the same age, who will be diagnosed in the time period under consideration. That is to say that if the 10-year risk of colon cancer for a 50-year-old woman is 17 then among 1000 women age 50, on average, 17 cases of colon cancer would be diagnosed.

## Interactions

Each exposure included in the Harvard Cancer Risk Index may have an interactive effect with other environmental, behavioral, occupational or lifestyle factors. The joint effects of these factors may substantially alter an individual's cancer risk. For example smoking, in combination with exposure to a known environmental carcinogen such as asbestos, can increase considerably the absolute risk of lung cancer, though this interaction may be less than multiplicative [20, 21]. Likewise, alcohol and tobacco smoking interact synergistically to increase risk of mouth and pharyngeal cancer [22]. The interactive effects of exposures other than asbestos were not accommodated in the Harvard Cancer Risk Index since their magnitude has not been precisely evaluated in most instances. Further, the added complexity of these possible interactions makes a pencil-and-paper risk index more difficult than can be justified based on their contribution to cancer risk.

### *Effects of exposures on other causes of morbidity*

Many exposures associated with cancer are also associated with other chronic and infectious diseases. For example, moderate physical activity reduces the risk of cancer as well as of diabetes, depression, coronary artery disease, and obesity [23]. Alcohol increases risk of breast cancer but reduces the risk of cardiovascular mortality [24]. A broader application would integrate the multiple benefits of lifestyle modification in reducing the total burden of all major diseases. Such additions will likely aid in focusing clinical counseling.

### *Effect of age*

The risk of being diagnosed with most cancers increases with age. However, the best approach to communicating the effect of age on risk is not clear. We have chosen to present the 10-year risk of cancer at the ages of 40, 50 and 60. Other ages could be added in a computerized version of the index.

One remaining concern with the focus of the Harvard Cancer Risk Index is that we do not address prevention efforts that might be acting earlier in life, such as avoiding sunburns to reduce the risk of melanoma [25]. While age is an important risk factor for cancer, the greatest opportunity to prevent some specific cancers is during early adolescence when an individual's risk profile is largely determined. Health protective habits such as physical activity, diets rich in fruits and vegetables, and choices about smoking and alcohol intake are often adopted during this time period [26]. Interventions targeting adolescence can help shape healthy lifestyles and prevent risk behaviors that may otherwise persist throughout the life course.

Research also indicates that age of exposure can differentially affect cancer risk. For example, the younger the age of exposure to ionizing radiation, the greater is the risk of subsequent breast cancer [27]. Likewise, earlier age at initiation of cigarette smoking is exponentially related to lung cancer risk [28]. These refinements of the details of exposure are not readily amenable to incorporation into the Harvard Cancer Risk Index. However, the population average age at exposure is reflected in the relative risk estimates. Thus for exposures with little range – such as age at starting to smoke, we are unlikely to have created substantial misclassification of risk.

### *Time frame for benefits from changes in lifestyle*

The time frame for behavior change to alter risk of cancer is not clearly defined. Any changes in lifestyle adopted as a consequence of perceived cancer risk will in

the long run reduce cancer risk, on average. Risk of lung cancer may be halved within 5 years of cessation from smoking [29, 30]. Reduction in cancer risk may, however, take 10 or more years before the full benefit of the changes in lifestyle are seen in the form of lower cancer risks. For example, regular aspirin use appears to require more than 10 years to demonstrate reduction in risk of colon cancer. Further, the rate of accumulating benefits may vary by cancer site and in conjunction with other factors. Thus the Harvard Cancer Risk Index should not be considered as a tool to identify high-risk individuals for preventive interventions, but rather as an aid to a population-wide shift in the prevalence of risk factors that will reduce the population burden of cancer.

### *Limitations*

The goal of the Harvard Cancer Risk Index is to quantify established and probable factors affecting cancer incidence. It should be considered a general guide for assessing an individual's risk of cancer. The scoring system does not allow precise estimation of an individual's risk of cancer. A constellation of personal, environmental, and behavioral factors, specific to the individual, may differentially affect his or her risk of cancer and cannot be accounted for in the Harvard Cancer Risk Index. This risk index represents an average and is based only on the factors accounted for in the model.

The estimates used in this risk index are based on the existing literature and the judgement of environmental, nutritional, and occupational health researchers of the Harvard Center for Cancer Prevention. These estimates do not necessarily represent the views of all members of the broader scientific community, although differences in opinion are unlikely to have major consequences in the estimates. One of the main reasons for using a categorical approach to risk classification was to reduce the likely disagreement among investigators.

We have limited the Harvard Cancer Risk Index to the leading causes of cancer. For women and men over age 50, these cancers account for over 90% of the cumulative risk of total cancer over the next 30 years, though only 80% of the total number of cases of cancer diagnosed in any 1 year. Therefore, though the Harvard Cancer Risk Index appears incomplete, it nevertheless should serve as a useful guide for counseling patients about their cancer risk.

Estimates by race/ethnicity are not readily included in a pencil-and-paper version of the risk index. Further refinement of the Harvard Cancer Risk Index using a computer-based presentation might incorporate all these distinctions. Overall cancer rates vary considerably for

black men and white men, and less so for women. Therefore, with a computer interface that allows for more complexity than a pencil-and-paper version, it will be possible to use race-specific probabilities of cancer incidence rather than those derived from the overall US rates for men and women. A computer interface will also provide the potential for interactive real-time feedback to the users regarding change in risk with behavior changes. This feedback and tailored messaging might improve counseling for behavior change [31]. Further research is needed to evaluate these potential features.

Because the cancer research base continues to expand, cancer risks from specific exposures will be continuously updated. Further evaluation of the Harvard Cancer Risk Index using data from other communities or countries could be a useful next step.

## Conclusions

The Harvard Cancer Risk Index offers a simple estimation of risk for cancer. It may help inform users of the major factors contributing to their risk, and of the changes that would reduce this risk. Presentation of data in several formats will aid users to understand the risk estimates. The Harvard Cancer Risk Index strictly applies to current cancer risk in the United States. Testing of the index against cancer risk in other countries will be required before its widespread use, although the underlying relations between lifestyle and cancer risk should apply to the majority of Western societies.

## Appendix

Members of the Risk Index Working Group include: K. A. Atwood, W. Y. Chen, D. Christiani, G. A. Colditz, W. DeJong, K. Emmons, A. L. Frazier, C. Fuchs, D. J. Hunter, R. R. Monson, B. J. Rockhill, L. Rhomberg, D. Trichopoulos, M. C. Weinstein, and W. C. Willett.

## Acknowledgements

The Harvard Cancer Risk Index was developed through the contribution of time by faculty and support from the operating budget of the Harvard Center for Cancer Prevention.

## References

1. Kannel WB, McGee D, Gordon T (1976) A general cardiovascular risk profile: the Framingham study. *Am J Cardiol* **38**: 46–51.
2. Kannel WB, McGee DL (1987) Composite scoring – methods and predictive validity: insights from the Framingham Study. *Health Serv Res* **22**: 501–535.
3. Morrow K, Morris CK, Froelicher VF, et al. (1993) Prediction of cardiovascular death in men undergoing noninvasive evaluation for coronary artery disease. *Ann Intern Med* **118**: 689–695.
4. Gail MH, Brinton LA, Byar DP, et al. (1989) Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. *J Natl Cancer Inst* **81**: 1879–1886.
5. Constantino J, Gail M, Pee D, et al. (1999) Validation studies for models projecting the risk of invasive and total breast cancer incidence. *J Natl Cancer Inst* **91**: 1541–1548.
6. Dupont WD, Plummer WD (1996) Understanding the relationship between relative and absolute risk. *Cancer* **77**: 2193–2199.
7. Colditz GA, DeJong W, Hunter DJ, Trichopoulos D, Willett WC, eds (1996) Harvard Report on Cancer Prevention. *Cancer Causes Control* **7**(Suppl.): S1–S55.
8. Black W, Nease R, Tosteson A (1995) Perception of breast cancer risk and screening effectiveness in women younger than 50 years of age. *J Natl Cancer Inst* **87**: 720–731.
9. US Preventive Services Task Force. (1996) *Guide to Clinical Preventive Services*. Baltimore, MD: Williams & Wilkins.
10. US Preventive Services Task Force. Screening for colorectal cancer. (1996) *Guide to Clinical Preventive Services*, 2nd edn. Baltimore, MD: Williams & Wilkins, pp. 89–103.
11. US Preventive Services Task Force. Screening for cervical cancer. (1996) *Guide to Clinical Preventive Services*, 2nd edn. Baltimore, MD: Williams & Wilkins, pp. 105–117.
12. International Agency for Research on Cancer (1987) *Overall Evaluations of Carcinogenicity: An Update of IARC Monographs Volumes 1 to 42*. Lyon, France: IARC.
13. Laupacis A, Sekar N, Stiell I (1997) Clinical prediction rules. A review and suggested modifications of methodological standards. *JAMA* **277**: 488–494.
14. Wasson J, Sox H, Neff R, Goldman L (1985) Clinical prediction rules: application and methodological standards. *N Engl J Med* **313**: 793–799.
15. Stiell I, Wells G, Laupacis A (1995) Multicentre trial to introduce the Ottawa ankle rules for use of radiography in acute ankle injury. *BMJ* **311**: 595–597.
16. Calman KC (1996) Cancer: science and society and the communication of risk. *BMJ* **313**: 799–802.
17. Calman KC, Royston GHD (1997) Risk language and dialects. *BMJ* **315**: 939–942.
18. Bunker J, Houghton J, Baum M (1998) Putting the risk of breast cancer in perspective. *BMJ* **317**: 1307–1309.
19. Chu K, Tarone R, Chow W, Hankey B, Ries L (1994) Temporal patterns in colorectal cancer incidence, survival, and mortality from 1950 through 1990. *J Natl Cancer Inst* **86**: 997–1006.
20. Steenland K, Thun M (1986) Interaction between tobacco smoking and occupational exposure in the causation of lung cancer. *J Occup Med* **43**: 111–118.
21. Berry G, Newhouse ML, Antonis P (1985) Combined effects of asbestos and smoking on mortality from lung cancer and mesothelioma. *Br J Ind Med* **42**: 12–18.
22. Rothman K, Keller A (1972) The effect of joint exposure to alcohol and tobacco on risk of cancer of the mouth and pharynx. *J Chron Dis* **23**: 711–716.
23. US Department of Health and Human Services (1996) *Physical Activity and Health*. Atlanta, GA: US Department of Health and

- Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion.
24. Thun M, Peto R, Lopez A, *et al.* (1997) Alcohol consumption and mortality among middle-aged and elderly US adults. *N Engl J Med* **337**: 1705–1714.
  25. English DR, Armstrong BK, Kricker A, Fleming C (1997) Sunlight and cancer. *Cancer Causes Control* **8**: 271–283.
  26. US Department of Health and Human Services (1994) *Report of the Surgeon General. Preventing Smoking Among Adolescents*. Washington, DC: Centers for Disease Control and Prevention, US DHHS. US Public Health Service.
  27. Tokunaga M, Land C, Tokuoka S, Nishimori I, Soda M, Akiba S (1994) Incidence of female breast cancer among atomic bomb survivors, 1950–1985. *Radiation Res* **138**: 209–223.
  28. Brown C, Chu K (1987) Use of multistage models to infer stage affected by carcinogenic exposure: example of lung cancer and cigarette smoking. *J Chron Dis* **40** (Suppl. 2): 171s–179s.
  29. US Department of Health and Human Services (1990) *The Health Benefits of Smoking Cessation*. Washington, DC: US Department of Health and Human Services, Public Health Service, Centers for Disease Control, Center for Chronic Disease Prevention and Health Promotion, Office of Smoking and Health, vol. DHSS Publication No (CDC)90-8416.
  30. Speizer F, *et al.* (1999) Prospective study of smoking, antioxidant intake, and lung cancer in middle-aged women. *Cancer Causes Control* **10**: 475–.
  31. Strecher V, Greenwood T, Wang C, Dumont D (1999) Interactive multimedia and risk communication. *Monogr Natl Cancer Inst* **25**: 134–139.