Cancer Epidemiology 34 (2010) 534–541

Contents lists available at ScienceDirect

Cancer Epidemiology

The International Journal of Cancer Epidemiology, Detection, and Prevention

journal homepage: www.cancerepidemiology.net

10-Year risk of colorectal cancer: Development and validation of a prediction model in middle-aged Japanese men

Enbo Ma, Shizuka Sasazuki *, Motoki Iwasaki, Norie Sawada, Manami Inoue

Shoichiro Tsugane for the Japan Public Health Center-based Prospective Study Group

Epidemiology and Prevention Division, Research Center for Cancer Prevention and Screening, National Cancer Center, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan

ARTICLE INFO

Article history:
Accepted 30 April 2010
Available online 31 May 2010

Keywords:
Prospective study
Epidemiologic method
Risk prediction
Colorectal cancer
Validation
Japanese men

ABSTRACT

Background: To estimate an individual's probability of developing colorectal cancer (CRC) may aid health professionals and individuals in improving lifestyle behaviors or deciding the screening regimens. As fewer studies on cancer risk prediction were seen so far, we initially developed an assessment tool with synthesizing key information from a variety of CRC risk factors through a large population-based cohort study. *Method:* The prediction model was derived from 28,115 men in the Japan Public Health Center-based (JPHC) Prospective Study Cohort II (follow-up: 1993–2005), with risk factors selected by Cox proportion hazard regression. 18,256 men in the JPHC Study Cohort I (follow-up: 1995–2005) were used to evaluate the model's performance. *Results:* 543 and 398 CRCs were diagnosed during the follow-up period in Cohorts II and I, respectively. The prediction model, including age, BMI, alcohol consumption, smoking status, and the daily physical activity level, showed modest discrimination ability for CRC ($C = 0.70$; 95% confidential interval, 0.68–0.72) in Cohort II and well calibrated in Cohort I (Hosmer–Lemeshow $\chi^2 = 14.2$, $P = 0.08$). *Conclusion:* The 10-year CRC risk prediction model may be used to estimate CRC risk in Japanese men. It may also play a role in the promotion of CRC prevention strategies.

1. Introduction

Colorectal cancer (CRC) was the second most commonly diagnosed cancer in the Japanese population in 2002 [1,2]. Approximately 11% of total cancer deaths in men and 14% in women were from CRCs in 2005 [2]. The high morbidity and mortality noted in the Japanese population were similar to those in North American and European counties [3].

Some risk factors for CRC were documented in the revised expert report from the World Cancer Research Fund, including physical activity, alcohol consumption, body and abdominal fatness, and consumption of vegetables and foods containing fiber [4]. A recent meta-analysis confirmed that smoking was significantly associated with CRC incidence and mortality [5]. In epidemiologic studies of the Japanese population, the risk factors of physical activity [6,7], alcohol consumption [8,9], smoking habit [8,9], and body mass index (BMI) [9,10] were consistently identified, whereas consumption of vegetables [11] and foods containing fiber [12] were not. Systematic reviews of large studies in Japan also verified the findings for alcohol consumption [13] and smoking habit [14]. In the Japanese population, however, these risk factors were more prevalent in men than in women, and little evidence of modifying CRC risk by reproductive factors has been found among Japanese women [15,16]. Nevertheless, most of these established risk factors for CRC are modifiable, and their improvement has been incorporated into primary cancer prevention strategies in Japan [17].

Given the high incidence of CRC and its significant cost to society, it is critical to reduce the identified risk factors in order to prevent CRC in a population. An individual's risk probability of developing CRC could be estimated by using information on established factors, which would aid physicians and individuals in improving lifestyle behavior and/or deciding on screening regimens for CRC prevention [17–19]. Moreover, from the public health point of view, risk prediction tools could also be used to effectively disseminate information on cancer prevention.

Several studies estimated the absolute risk probability of developing CRC, although they were based on case–control study [18], expert opinion [20], or specific populations [21,22]. In this paper, we present a CRC risk prediction model in Japanese men, derived and validated by two large cohorts from the Japan Public Health Center-based (JPHC) Prospective Study. We also present a simplified score model that can be easily used to estimate an individual's absolute CRC risk based on lifestyle information.

* Corresponding author. Tel.: +81 3 3542 2511x3378; fax: +81 3 3547 8578.
E-mail address: ssasazuk@ncc.go.jp (S. Sasazuki).

1877-7821/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.canep.2010.04.021

2. Materials and methods

2.1. Study participants

In the JPHC Study, Cohort I, with participants aged 40–59 years, was launched in 1990 and Cohort II, with participants aged 40–69 years, was added in 1993. A total of 48,448 men were initially identified in 11 public health center-based (PHC) areas throughout Japan. The details of the study design and baseline response have been described elsewhere [23,24]. The study was approved by the Institute Review Board of the National Cancer Center, Tokyo, Japan.

The baseline survey for Cohort II had more comprehensive data on physical activity and the food frequency questionnaire (FFQ) (52 food items) than those and the FFQ (44 food items) for Cohort I. In the 5-year follow-up survey, all investigations including the FFQ (138 food items) were the same for both cohorts. Considering the inconsistency of questionnaires and follow-up periods of the two cohorts, in the present study we used the baseline survey of Cohort II men to derive the risk prediction model of CRC and the 5-year follow-up survey of Cohort I men to validate the model.

Participants who reported a history of cancer or cardiovascular disease, were diagnosed with cancers, or were censored before the start of the follow-up survey were excluded, leaving 28,115 eligible subjects for model derivation in Cohort II and 18,256 for model validation in Cohort I.

2.2. Risk factor measurements

Self-administered questionnaires contained items on demographic characteristics, medical history, smoking habit, alcohol consumption, physical activity, occupation, and other factors, as well as diets by validated FFQs [25,26].

BMI was calculated as weight in kilograms divided by the square of height in meters. Physical activity levels, measured by metabolic equivalent (MET) hours per day, were estimated by multiplying the reported time spent at each activity per day by its assigned MET intensity: heavy physical work or strenuous exercise (4.5), walking or standing (2.0), sedentary (1.5), and sleep or others (0.9) [6,27]. Daily physical activity level was the sum of MET-hour scores across all activities.

Smoking habit was grouped into never, former, and current smokers. Alcohol consumption was categorized into four groups (never, occasional, regular $<$300 g/week, and regular $\geq$300 g/week), in which regular drinkers were categorized by multiplying the frequency per week by the usual daily amount of alcohol consumed [8].

Daily food intake was calculated by multiplying the frequency by standard portion size and relative size for each food item in the FFQ. Daily intake of nutrients was calculated using the 5th revised edition of the Standard Tables of Food Composition in Japan [28].

2.3. Follow-up and case assessment

Participants were followed until 31 December 2005. Residence status, movement of households, and survival were confirmed annually using the residential registers. Information on the cause of death was obtained by examining the death certificates provided by the Ministry of Health, Labour, and Welfare. The occurrence of cancer was identified by active patient notification through the major local hospitals in the study areas and data linkage with population-based cancer registries. The site and histology of each cancer were coded using the International Classification of Diseases for Oncology, 3rd edition (ICD-O-3), with C18–C20 for CRC, C180–C189 for colon cancer, and C199 and C209 for rectal cancer.

2.4. Statistical analysis

Person-years of follow-up were counted from the date of survey response (1993 for Cohort II and 1995 for Cohort I) until the date of CRC diagnosis, the date of moving out of a study area, the date of death, or the end of 2005, whichever came first. Persons lost to follow-up were censored on the last confirmed date of their presence in the study area. Extreme values of height ($<$100 or $>$199 cm), weight ($<$20 kg), and BMI ($<$14 or $>$40 kg/m^2) were removed from this analysis. Nutrient intakes were categorized into tertiles for all study participants, with the lower tertile as the reference.

2.4.1. Prediction model derived by JPHC Cohort II

Cox proportional hazards models were derived after testing for the assumptions underlying its use. Then the model of predictive risk of developing CRC was fitted, in which the average survival rates at follow-up time points were estimated by baseline hazard function with mean values of potential predictors. Hazard ratios (HR) and 95% confidential interval (CI) of each risk factor were also estimated. Based on the previous publications in Japanese populations and age-adjusted univariate analysis performed for available variables in this study (including more than 30 food items and nutrients), the potential predictors were applied for building the full multivariate model, which including age, BMI, daily physical activity, alcohol consumption, smoking habit, family history of CRC, and diabetes diagnosed, and interested interaction terms with biological plausibility between alcohol and smoking, and physical activity and BMI. PHC areas were treated as strata in the analysis; assessment of likely shrinkage (over-fitting) was evaluated for the reduced models by $[LR-(p-q)-q]/[LR-(p-q)]$, where LR denotes the likelihood ratio χ^2, and p and q denote the regression degrees of freedom for the full model and for a reduced model, respectively [29]. Non-linear relationships (transformations) of age, BMI, or daily physical activity were tested by using multiple fractional polynomial method of two degree [30,31], however, none of which had been statistically significant for leaving in the model.

For each risk factor, the regression coefficients of two cohorts were compared by a 2-tailed Z statistics, $Z=(\beta_{[d]}-\beta_{[v]})/SE$, where $\beta_{[d]}$ and $\beta_{[v]}$ are the regression coefficients of Cohort II and Cohort I, respectively, and SE is the standard error of the difference in the coefficients, calculated as $\sqrt{(SE^2_{\beta_{[d]}}\pm SE^2_{\beta_{[v]}})}$ [32]. The Z statistic was used to test the difference in HR of each risk factor/category between the two cohorts [32]. The individual risk of CRC was estimated based on the baseline hazard function of the Cox regression model derived from Cohort II, which method was same as one developed in Framingham heart study [33], where $P=1-S(t)^{exp(f[x,M])}$ and $f(x,M)=\beta 1(x1-M1)+\ldots+\beta j(xj-Mj)$. $\beta 1,\ldots,\beta j$ are the regression coefficients, $x1,\ldots,xj$ represent an individual's risk factors, $M1,\ldots,Mj$ are the mean values of the risk factors in the cohort (for category variables, $x1,\ldots,xj$ are the dichotomous value of the created dummy variable for each category, entering 1 if the individual's value fits that certain category and 0 otherwise, and $M1,\ldots,Mj$ are the proportion of the certain category of the variable in the cohort), and $S(t)$ is the average survival rate at time t of subjects with the mean values of the risk factors used in the Cox model. This procedure performed a better validity than prepared by Ederer method [34]. The predicted 10-year risk of CRC, therefore, was estimated by the baseline hazard function of Cohort II with mean values of each predictor at the 10-year follow-up time.

2.4.2. Prediction model validated by JPHC Cohort I

Discrimination, the ability of a predictive model to separate those who experience an event from those who do not, was

Table 1
Full and reduced predicative models for estimation of developing colorectal cancer events in Cohort II men, Japan Public Health Center-based Prospective Study, 1993–2005.

Variables retained	Full model			Reduced 1[a]			Reduced 2[b]		
	β	S.E.(β)	*P*-Value	β	S.E.(β)	*P*-Value	β	S.E.(β)	*P*-Value
CRC[c]									
Age, year	0.079	0.006	<0001	0.080	0.006	<0001	0.080	0.006	<0001
BMI, kg/m^2	0.001	0.061	0.98	0.047	0.016	<0.01	0.047	0.016	<0.01
Physical activity, MET-h/d	−0.055	0.049	0.27	−0.019	0.006	<0.01	−0.019	0.006	O.01
Family history of CRC (yes)	−0.085	0.382	0.82	−0.087	0.382	0.82	–	–	–
Diabetes (yes)	0.103	0.160	0.52	0.095	0.160	0.55	–	–	–
Alcohol consumption[d]									
Never	0.052	0.244	0.83	−0.163	0.210	0.44	−0.163	0.210	0.44
Regular (<300 g/w)	0.393	0.230	0.09	0.359	0.192	0.06	0.358	0.192	0.06
Regular (≧300 g/w)	0.584	0.273	0.03	0.657	0.195	O.001	0.659	0.195	O.001
Smoking									
Former	−0.165	0.196	0.40	0.070	0.133	0.60	0.071	0.133	0.59
Current	−0.225	0.330	0.50	0.237	0.119	0.05	0.239	0.119	0.04
Smoking × alcohol	0.078	0.056	0.17	–	–	–	–	–	–
BMI × physical activity	0.002	0.002	0.46	–	–	–	–	–	–
d.f.	12			10			8		
Likelihood ratio x^2	239.8			237.3			241.2		
Shrinkage	–			0.96			0.97		
C-Index	0.703			0.699			0.699		
Colon cancer									
Age, year	0.084	0.008	<0001	0.085	0.008	<0001	0.085	0.008	<0001
BMI, kg/m^2	0.037	0.079	0.64	0.048	0.021	0.02	0.049	0.021	0.02
Physical activity, MET-h/d	−0.028	0.063	0.66	−0.019	0.008	0.02	−0.020	0.008	0.01
Family history of CRC (yes)	0.438	0.384	0.25	0.437	0.384	0.26	–	–	–
Diabetes (yes)	0.330	0.188	0.08	0.323	0.188	0.09	–	–	–
Alcohol consumption[d]									
Never	0.077	0.323	0.81	−0.133	0.276	0.63	−0.140	0.276	0.61
Regular (<300 g/w)	0.493	0.305	0.11	0.431	0.253	0.09	0.419	0.254	0.10
Regular (≧300 g/w)	0.651	0.363	0.07	0.657	0.257	0.01	0.655	0.258	0.01
Smoking									
Former	−0.006	0.258	0.98	0.180	0.173	0.30	0.186	0.173	0.28
Current	−0.012	0.433	0.98	0.341	0.157	0.03	0.347	0.157	0.03
Smoking × alcohol	0.057	0.073	0.44	–	–	–	–	–	–
BMI × physical activity	0.000	0.003	0.90	–	–	–	–	–	–
d.f.	12			10			8		
Likelihood ratio x^2	165.7			165.1			166.0		
Shrinkage	–			0.94			0.95		
C-Index	0.710			0.710			0.708		
Rectal cancer									
Age, year	0.072	0.009	<0001	0.071	0.009	<0001	0.067	0.009	<0001
BMI, kg/m^2	−0.054	0.098	0.58	0.033	0.025	0.19	–	–	–
Physical activity, MET-h/d	−0.097	0.078	0.22	−0.018	0.010	0.07	−0.020	0.008	0.02
Diabetes (yes)	−0.357	0.311	0.25	−0.078	0.240	0.75	–	–	–
Alcohol consumption[d]									
Never	0.027	0.374	0.94	−0.401	0.291	0.17	−0.094	0.361	0.80
Regular (<300 g/w)	0.261	0.349	0.45	0.083	0.259	0.75	0.365	0.335	0.28
Regular (≧300 g/w)	−0.536	0.514	0.30	0.488	0.268	0.07	0.745	0.281	0.01
Smoking									
Former	−0.3%	0.305	0.19	0.088	0.181	0.63	–	–	–
Current	0.504	0.415	0.22	0.088	0.181	0.63	–	–	–
Smoking × alcohol	0.109	0.087	0.21	–	–	–	–	–	–
BMI × physical activity	0.003	0.003	0.31	–	–	–	–	–	–
d.f.	11			9			5		
Likelihood ratio χ^2	82.9			80.0			75.7		
Shrinkage	–			0.89			0.94		
C-Index	0.698			0.678			0.678		

[a] Removed interactions.
[b] Further removed family history and diabetes diagnosed for CRC and colon cancer; diabetes diagnosed, BMI, and smoking habit for rectal cancer.
[c] CRC, colorectal cancer; MET, metabolic equivalent.
[d] Occasional alcohol consumption was as the reference.

assessed using the *C* statistic, the area under the receiver operating characteristic curve [32]. The overall *C* statistics and its 95% CIs were calculated by logistic regressions. Calibration is another measure of performance of a prediction model that tests how closely predicted outcomes agree with actual outcomes [32,35]. The calibration was conducted in Cohort I, using the β coefficients, the mean of each risk factor, and the average survival rate at 10-year from the original Cohort II. Participants in Cohort I were divided into 10 deciles of individual predicted risk, and in each decile the expected events were the sum of individual predicted

Table 2
Characteristics of risk factors, person-years of follow-up, and colorectal cancer events in men, Japan Public Health Center-based Prospective Study, 1993–2005[a].

Risk factor	Cohort II[b]						Cohort I[c]					
	Participants, mean (SD), %	No. of participants	Person-years of follow-up	No. of events			Participants, mean (SD), %	No. of participants	Person-years of follow-up	No. of events		
				CRC	Colon	Rectum				CRC	Colon	Rectum
Age, year	52.9(8.8)	28,115	310,059	543	329	214	54.7 (6.0)	18,256	184,496	389	239	150
BMI, kg/m^2	23.4 (2.9)	28,115	310,059	543	329	214	23.6 (2.8)	18,256	184,496	389	239	150
Physical activity, MET-h/d	28.7(7.3)	27,284	300,982	523	314	209	26.8 (7.0)	17,112	173,159	361	219	142
Alcohol consumption												
Never	23.5	6,355	68,967	96	60	36	23.2	4,192	41,652	83	51	32
Occasional	7.7	2,087	23,652	26	15	11	8.6	1,565	16,013	22	10	12
Regular: <300 g/w	48.1	13,038	143,999	248	155	93	35.4	6,403	65,130	108	64	44
Regular: ≧300 g/w	20.8	5,623	62,184	146	85	61	32.9	5,948	60,187	171	111	60
Smoking status												
Never	23.6	6,579	74,342	111	64	47	36.1	6,483	66,178	110	68	42
Former	23.9	6,657	73,238	142	89	53	16.2	2,901	29,256	78	57	21
Current	52.5	14,601	159,481	284	174	110	47.7	8,555	85,836	195	112	83

[a] CRC, colorectal cancer; MET, metabolic equivalent.
[b] Cohort II (follow-up: 1993–2005) was used to develop the prediction model.
[c] Cohort I (follow-up: 1995–2005) was to evaluate the prediction model's performance.

risk [36]. The Hosmer–Lemeshow χ^2 test was applied to analyze the difference between the observed and estimated risk by groups of deciles [37]. The ratio of observed and expected CRC events (the sum of individual predicted risk probability in a certain risk category) was used to test the model predictive capability for each risk factor in Cohort I. The 95% CIs for O/E ratio was calculated as $(O/E) \times \exp[\pm 1.96\sqrt{(1/O)}]$; the prediction model underestimated the CRC risk if the O/E ratio was >1, while it overestimated the risk if the O/E ratio was <1 [36].

2.4.3. Simple point score model

A simple point score model (risk sheet) for CRC was developed based on the original prediction model, with the transference of continuous variables of age, BMI, and physical activity into category variables [38,39]. The β coefficients were newly fitted by the Cox model with each of category variables. The first step was to round regression coefficients to scores, and in this analysis, we multiplied coefficients by three, and round them [38,40]. Further, the risk score of each participant was assigned by summing the points from each risk factor present. The score sheets provide comparison 10-year absolute risks for persons of the same age from average and low-risk CRC.

All analyses were conducted using SAS version 9.01 (SAS Inc., Cary, NC, USA).

3. Results

As of December 2005, newly diagnosed cases of CRC were 543 in Cohort II and 389 in Cohort I. In total, 310,059 and 184,496 person-years were observed in the average follow-up periods of 11.0 and of 10.1 years in Cohorts II and I, respectively.

Comparisons of model constructions among the full predictive model and the models with reduced variables were shown in Table 1, in which the reduced multivariate model with age, BMI, physical activity, smoking habit and alcohol consumption was the optimal one (the global test for model non-proportionality, $P = 0.984$, 0.597, and 0.093 for CRC, colon, and rectal cancer, respectively). Numbers of participants, person-years of follow-up, and CRC events, as well as the risk factors of CRC are listed in Table 2. The respective β coefficients and HRs for CRC risk factors obtained from Cox regression of Cohorts II and I, with baseline survival rate at 10-years, are shown in Table 3. Risk factors showed similar relationships to CRC, colon, and rectal cancer.

In the discriminatory analysis of Cohort II, the C statistics were 0.70 (95% CI, 0.68–0.72) for CRC, 0.71 (95% CI, 0.68–0.74) for colon cancer, and 0.68 (95% CI, 0.64–0.71) for rectal cancer, showing a good ability to distinguish cases from non-cases. In Cohort I, the C statistics were 0.64 (95% CI, 0.61–0.67) for CRC, 0.66 (95% CI: 0.62–0.70) for colon cancer, and 0.62 (95% CI: 0.57–0.66) for rectal cancer, showing a modest ability to distinguish cases from non-cases.

In the calibration analysis, χ^2 was 14.2 ($P = 0.08$) for CRC, 11.0 ($P = 0.20$) for colon, and 11.2 ($P = 0.19$) for rectum cancer, showing that the actual rates of CRC in Cohort I were similar to the rates predicted by the Cohort II function (Fig. 1). The overall O/E ratios were 1.09 (95% CI, 0.98–1.23) for CRC, 1.19 (95% CI, 1.03–1.37) for colon cancer, and 0.94 (95% CI, 0.78–1.12) for rectal cancer. Agreement between the predicted and the observed number of events was good in most risk factor categories with several exceptions (e.g., underestimation for CRC in the "never" alcohol consumption category and overestimation for rectal cancer in the age group of 45–49) (Table 4).

In addition, when participants who had a history of diabetes (1991 in Cohort II and 1332 in Cohort I) or a family history of CRC in first-degree relatives (475 in Cohort II and 157 in Cohort I) were excluded, the same predictive risk factors were identified, and similar discrimination and calibration values were observed for CRC, colon, and rectal cancer, respectively, in Cohort I (data not shown).

The simple point score model (risk sheet) was developed for CRC in Cohort II (Fig. 2), for which the C statistic was 0.69 (95% CI, 0.67–0.71). In Fig. 2, the average and the lowest risk probability by age groups in Cohort II are also shown. Correspondingly, validation was performed in Cohort I for the simple point score model: the C statistic was 0.61 (95% CI, 0.58–0.64) for CRC, with similar O/E ratios and 95% CIs in each category of risk factors (data not shown).

4. Discussion

We developed a CRC risk prediction model with established risk factors of age, BMI, alcohol consumption, smoking status, and physical activity level for middle-aged Japanese men. The prediction model was well calibrated in an external cohort. We also presented a simple point score model (risk sheet) for CRC risk estimation.

Cancer is a multifactorial disease involving a variety of factors in the development of clinical manifestations. This recognition has

Table 3
β-Coefficients and hazard ratios with 95% confidence intervals of colorectal cancer risk factors in men, Japan Public Health Center-based Prospective Study, 1993–2005[a].

Risk factor	Cohort II[b,d]						Cohort I[c,d]					
	CRC		Colon		Rectum		CRC		Colon		Rectum	
	β	HR (95% CI)	β	HR (95% CI)	β	HR (95%CI)	β	HR (95% CI)	β	HR (95% CI)	β	HR (95% CI)
Age, year	0.080	1.08 (1.07–1.10)	0.085	1.09 (1.07–1.11)	0.067	1.07 (1.05–1.09)	0.063	1.07 (1.05–1.09)	0.062	1.06 (1.04–1.09)	0.065	1.07 (1.04–1.10)
BMI, kg/m^2	0.047	1.05 (1.02–1.08)	0.049	1.05 (1.01–1.09)	–	–	0.003	1.00 (0.97–1.04)	0.013	1.01 (0.97–1.06)	–	–
Physical activity, MET-h/d	−0.019	0.98 (0.97–0.99)	−0.020	0.98 (0.97–1.00)	−0.020	0.98 (0.97–1.00)	−0.017	0.98 (0.97–1.00)	−0.027	0.97 (0.95–0.99)	−0.006	0.99 (0.97–1.02)
Alcohol consumption												
Never	−0.163	0.85 (0.56–1.28)	−0.140	0.87 (0.51–1.49)	−0.149	0.86 (0.48–1.55)	0.314	1.37 (0.88–2.14)	0.474	1.61 (0.88–2.92)	−0.028	0.97 (0.51–1.87)
Occasional		1.00		1.00		1.00		1.00		1.00		1.00
Regular: <300 g/w	0.358	1.43 (0.98–2.09)	0.419	1.52 (0.93–2.50)	0.309	1.36 (0.80–2.31)	0.072	1.07 (0.69–1.67)	0.182	1.20 (0.67–2.16)	−0.197	0.82 (0.43–1.55)
Regular: ≧300 g/w	0.659	1.93 (1.32–2.83)	0.655	1.93 (1.16–3.19)	0.745	2.11 (1.21–3.65)	0.679	1.97 (1.30–3.00)	0.858	2.36 (1.35–4.14)	0.348	1.42 (0.76–2.63)
Smoking status												
Never		1.00		1.00				1.00		1.00		
Former	0.071	1.07 (0.83–1.39)	0.186	1.21 (0.86–1.69)	–	–	0.438	1.55 (1.15–2.09)	0.605	1.83 (1.27–2.64)	–	–
Current	0.239	1.27 (1.01–1.60)	0.347	1.41 (1.04–1.92)	–	–	0.323	1.38 (1.08–1.77)	0.222	1.25 (0.91–1.72)	–	–
Baseline survival function at 10-year, St(10)	0.9882		0.9928		0.9954		0.9835		0.9890		0.9942	

[a] CRC, colorectal cancer; HR, hazard ratio; CI, confidential interval; MET, metabolic equivalent.
[b] Cohort II (follow-up: 1993–2005) was used to develop the prediction model.
[c] Cohort I (follow-up: 1995–2005) was to evaluate the prediction model's performance.
[d] The HR of each risk factor/category was not significantly different between Cohort II and Cohort I ($P > 0.05$) for the model of CRC, colon, and rectal cancer, respectively.

Fig. 1. The 10-year observed and predicted colorectal cancer events in Cohort I men, Japan Public Health Center-based Prospective Study, 1993–2005.

led the development of risk assessment tools that attempt to synthesize the values of numerous variables into a single statement about the risk of developing a cancer [41]. In this prediction model, age, alcohol consumption, and daily physical activity level were identified as the most important CRC risk factors, consistent with other reports [4,18,20]. Although body weight was also a potential predictor in this analysis, BMI was arbitrarily selected in the model building as a relevant comprehensive risk factor of CRC [10,18,20].

Dietary factors such as consumption of red meat, green vegetables, fibers, dairy, calcium supplement use, or intake of folate were not identified in this population, although they were previously reported as possibly related to CRC risk [4,18,42]. Moreover, no dietary food combinations, including total meat (pork, beef, bacon, ham, and sausage) [42], processed meat (bacon, ham, and sausage) [42,43], total white meat (fish and poultry) [42], ratio of red meat to vegetable, or ratio of red meat to white meat [44] were risk predictors of CRC in this study population. Although in recent years the dietary pattern in the Japanese population has tended toward the western pattern, the traditional dietary habits were substantially maintained, especially in older people [45]. This may account for the lack of foods or dietary nutrients serving as significant factors for predicting CRC in men. Alternatively, it might be possible that data in this study population were insufficient to support a quantitative statement about the exact magnitude of risk from these diets.

A previous CRC risk prediction model was developed by means of larger case–control studies and included CRC screening during the previous 3 years and number of relatives with CRC [18]. In our study, sigmoidoscopy/colonoscopy and fecal occult blood test were not available in the Cohort II questionnaire, although these are known as indicators for the secondary prevention for CRC [46]. The personal history of diabetes was reported as a possible risk factor of CRC [26]. In the present study, however, diabetes showed statistical significance for colon cancer in the univariate analysis but not in the multivariate analysis. In addition, few participants reported a family history of CRC, such that this factor could not be considered for entering into the prediction model. In the analysis for participants without history of diabetes or family history of CRC, a similar predictive ability for CRC was observed. This may indicate that these two factors were not powerful enough for prediction of CRC in this population. Nevertheless, most CRC risk factors included in this prediction model represent lifestyle choices that can be modified with the aim of preventing the disease.

Several validation studies on cancer risk prediction models also showed modest discriminatory accuracy as measured by *C*

Table 4
10-Years of observed and expected colorectal cancer events, ratios and 95% confidential intervals in Cohort I men, Japan Public Health Center-based Prospective Study, 1993–2005[a].

	CRC					Colon					Rectum				
	Observed	Expected	*O*/*E* ratio	95%	CI	Observed	Expected	*O*/*E* ratio	95	%CI	Observed	Expected	*O*/*E* ratio	95%	CI
Overall	322	294	1.09	0.98	1.23	215	181	1.19	1.03	1.37	107	114	0.94	0.78	1.12
Age, years															
45–49	45	39.0	1.15	0.84	1.58	35	22.8	1.53	1.02	2.31	10	16.4	0.61	0.38	0.99
50–54	62	53.2	1.17	0.89	1.53	41	31.8	1.29	0.91	1.82	21	21.4	0.98	0.64	1.50
55–59	95	76.1	1.25	1.00	1.56	55	46.7	1.18	0.88	1.57	40	29.5	1.36	0.95	1.95
60–64	112	119.9	0.93	0.78	1.12	78	75.9	1.03	0.82	1.29	34	44.7	0.76	0.57	1.02
65–69	8	6.2	1.30	0.59	2.86	6	4.0	1.52	0.57	4.07	2	2.3	0.87	0.24	3.14
BMI, kg/m^2															
<25	230	200.9	1.14	1.00	1.31	153	123.6	1.24	1.04	1.48	–	–	–	–	–
≧25	92	93.5	0.98	0.80	1.21	62	57.6	1.08	0.83	1.39	–	–	–	–	–
Physical activity, MET-h/d															
<22.0	118	109.3	1.08	0.89	1.30	92	67.8	1.36	1.07	1.72	33	41.9	0.79	0.58	1.07
22.0-<28.9	95	101.4	0.94	0.77	1.14	70	62.4	1.12	0.87	1.44	34	39.4	0.86	0.63	1.18
≧28.9	83	83.6	0.99	0.80	1.23	57	50.9	1.12	0.85	1.47	33	33.1	1.00	0.71	1.40
Alcohol consumption															
Never	66	42.5	1.55	1.15	2.10	48	26.0	1.84	1.26	2.71	18	16.5	1.09	0.67	1.77
Occasional	19	17.7	1.07	0.67	1.71	9	10.6	0.85	0.47	1.56	10	6.4	1.57	0.72	3.42
Regular: <300 g/w	95	103.0	0.92	0.76	1.12	59	65.5	0.90	0.71	1.15	36	37.9	0.95	0.69	1.31
Regular: ≧300 g/w	137	129.6	1.06	0.89	1.26	96	78.2	1.23	0.98	1.53	41	53.1	0.77	0.59	1.01
Smoking status															
Never	87	91.6	0.95	0.77	1.17	58	52.7	1.10	0.84	1.44	–	–	–	–	–
Former	69	48.9	1.41	1.07	1.87	52	31.5	1.65	1.16	2.34	–	–	–	–	–
Current	160	149.7	1.07	0.91	1.25	103	94.6	1.09	0.89	1.33	–	–	–	–	–

[a] CRC, colorectal cancer; *O*/*E*, observed/expected; CI, confidential interval; MET, metabolic equivalent.

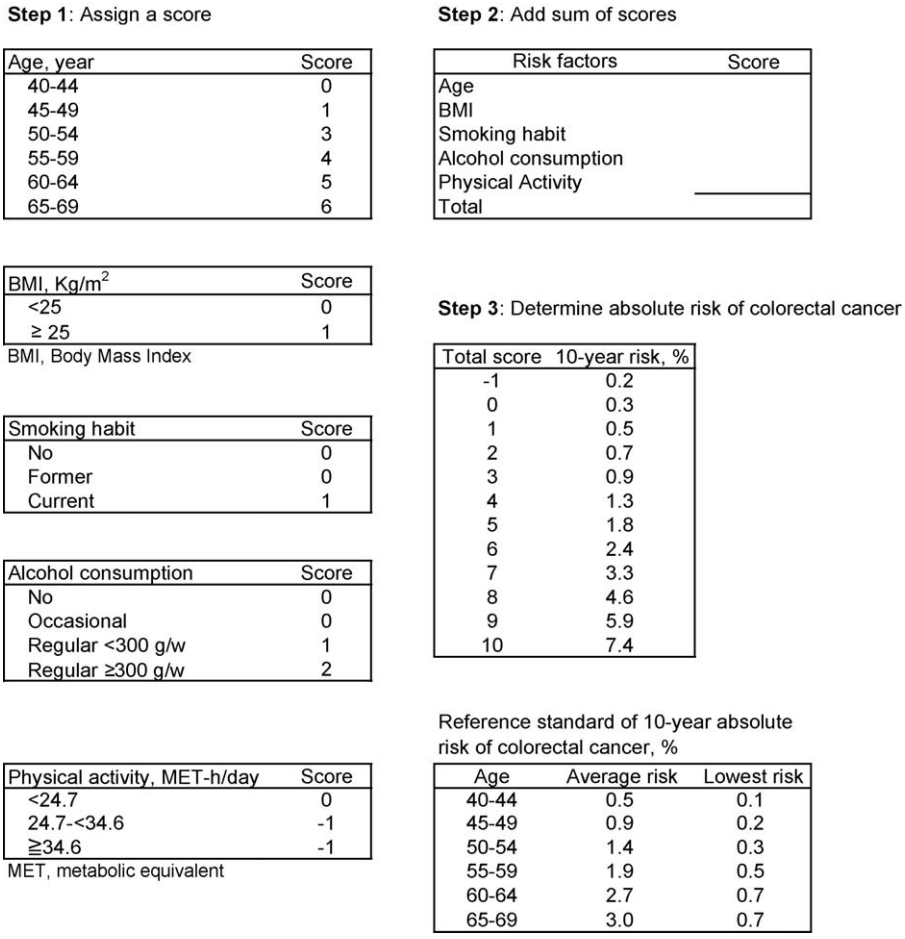

Fig. 2. Simple point score model (risk sheet) for evaluation of 10-year risk of colorectal cancer incidence in men.

statistics, including 0.61 for CRC [36], 0.60–0.63 for breast cancer [47,48], and 0.60–0.69 for lung cancer [49,50]. Similarly, the modest ability to predict CRC in this study suggested that in future studies stronger risk predictors need to be found [18], for instance, dietary nutrient intake or genotypes.

The overall predicted number of CRC events was close to the actual number, with several exceptions in the validation. The differences between the observed and the predicted CRC events in Cohort I may be due to a different distribution of participants with higher risk in the two cohorts. For example, more elderly men and smokers were in Cohort II than in Cohort I, while more heavy alcohol drinkers were in Cohort I than in Cohort II. The discrepancies in the questionnaires used in the two cohorts also may partly account for the difference [36].

The validation in this study was done in an external cohort (Cohort I); however, risk factor profiles and measurement were similar to those of the population for model development (Cohort II). Therefore, the generalizability of the prediction model needs to be tested in other populations to provide more external validations. Another limitation of this study was that the simple point score model (risk sheet) for estimation of CRC risk included not only simple frequency components (age, body weight, and smoking) but also those based on calculation (alcohol consumption by gram per week and physical activity by MET-hour per day). This may make it inconvenient for an individual to use the sheet directly. In addition, because the 5-year follow-up measurement was used as the baseline for Cohort I in this analysis, the smaller relevant population might reduce its validation capability.

In summary, the CRC risk prediction model was developed based on a large cohort study; it showed modest discrimination power and was well calibrated in another large cohort. This model may be used by clinicians, public health professionals, and individuals to estimate the CRC risk for Japanese men, which could play a role in the promotion of CRC prevention strategies. Further validation in other populations, with the addition of more established factors, is necessary.

Conflict of interest statement

None declared.

Acknowledgments

This study was supported by a Grant-in-Aid for Cancer Research (19 shi-2) and for the Third Term Comprehensive 10-year Strategy for Cancer Control (H21-Sanjigan-Ippan-003) from the Ministry of Health, Labour and Welfare of Japan and Grants-in-Aid for Scientific Research for Young Scientists (A) (19689014) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan and Japan Society for the Promotion of Science.

We thank all staff members in each study area for their painstaking efforts to conduct the baseline survey and follow-up. Members of the JPHC Study Group (principal investigator: S. Tsugane) include: S. Tsugane, M. Inoue, T. Sobue, and T. Hanaoka, Research Center for Cancer Prevention and Screening, National Cancer Center, Tokyo; J. Ogata, S. Baba, T. Mannami, A. Okayama, and Y. Kokubo, National Cardiovascular Center, Suita; K. Miyakawa, F. Saito, A. Koizumi, Y. Sano, I. Hashimoto, T. Ikuta, and Y. Tanaba, Iwate Prefectural Ninohe Public Health Center, Ninohe; Y. Miyajima, N. Suzuki, S. Nagasawa, Y. Furusugi, and N. Nagai, Akita Prefectural Yokote Public Health Center, Yokote; H. Sanada, Y. Hatayama, F. Kobayashi, H. Uchino, Y. Shirai, T. Kondo, R. Sasaki, Y. Watanabe, Y. Miyagawa, Y. Kobayashi, and M. Machida, Nagano Prefectural Saku Public Health Center, Saku; Y. Kishimoto, E. Takara, T. Fukuyama, M. Kinjo, M. Irei, and H. Sakiyama, Okinawa Prefectural Chubu Public Health Center, Okinawa; K. Imoto, H. Yazawa, T. Seo, A. Seiko, F. Ito, F. Shoji, and R. Saito, Katsushika Public Health Center, Tokyo; A. Murata, K. Minato, K. Motegi, and T. Fujieda, Ibaraki Prefectural Mito Public Health Center, Mito; T. Abe, M. Katagiri, M. Suzuki, and K. Matsui, Niigata Prefectural Kashiwazaki and Nagaoka Public Health Center, Kashiwazaki and Nagaoka; M. Doi, A. Terao, Y. Ishikawa, and T. Tagami, Kochi Prefectural Chuo-higashi Public Health Center, Tosayamada; H. Doi, M. Urata, N. Okamoto, F. Ide, and H. Sueta, Nagasaki Prefectural Kamigoto Public Health Center, Arikawa; H. Sakiyama, N. Onga, H. Takaesu, and M. Uehara, Okinawa Prefectural Miyako Public Health Center, Hirara; F. Horii, I. Asano, H. Yamaguchi, K. Aoki, S. Maruyama, M. Ichii, and M. Takano, Osaka Prefectural Suita Public Health Center, Suita; S. Matsushima and S. Natsukawa, Saku General Hospital, Usuda; M. Akabane, Tokyo University of Agriculture, Tokyo; M. Konishi, K. Okada, and I. Saito, Ehime University, Toon; H. Iso, Osaka University, Suita; Y. Honda, K. Yamagishi, S. Sakurai, and N. Tsuchiya, Tsukuba University, Tsukuba; H. Sugimura, Hamamatsu University, Hamamatsu; Y. Tsubono, Tohoku University, Sendai; M. Kabuto, National Institute for Environmental Studies, Tsukuba; S. Tominaga, Aichi Cancer Center Research Institute, Nagoya; M. Iida, W. Ajiki, and A. Ioka, Osaka Medical Center for Cancer and Cardiovascular Disease, Osaka; S. Sato, Osaka Medical Center for Health Science and Promotion, Osaka; N. Yasuda, Kochi University, Nankoku; K. Nakamura, Niigata University, Niigata; S. Kono, Kyushu University, Fukuoka; K. Suzuki, Research Institute for Brain and Blood Vessels Akita, Akita; Y. Takashima and M. Yoshida, Kyorin University, Mitaka; E. Maruyama, Kobe University, Kobe; M. Yamaguchi, Y. Matsumura, S. Sasaki, and S. Watanabe, National Institute of Health and Nutrition, Tokyo; T. Kadowaki, Tokyo University, Tokyo; M. Noda and T. Mizoue, International Medical Center of Japan, Tokyo; Y. Kawaguchi, Tokyo Medical and Dental University, Tokyo; H. Shimizu, Sakihae Institute, Gifu.

References

[1] Matsuda T, Marugame T, Kamo K, Katanoda K, Ajiki W, Sobue T. Cancer incidence and incidence rates in Japan in 2002: based on data from 11 population-based cancer registries. Jpn J Clin Oncol 2008;38:641–8.
[2] The Editorial Board of the Cancer Statistics in Japan, ed. Cancer statistics in Japan 2007. Tokyo: Foundation for Promotion of Cancer Research (FPCR), 2007
[3] Shibuya K, Mathers CD, Boschi-Pinto C, Lopez AD, Murray CJ. Global and regional estimates of cancer mortality and incidence by site. II. Results for the global burden of disease 2000. BMC Cancer 2002;2:37.
[4] Wiseman M. The second World Cancer Research Fund/American Institute for Cancer Research expert report. Food, nutrition, physical activity, and the prevention of cancer: a global perspective. Proc Nutr Soc 2008;67:253–6.
[5] Botteri E, Iodice S, Bagnardi V, Raimondi S, Lowenfels AB, Maisonneuve P. Smoking and colorectal cancer: a meta-analysis. JAMA 2008;300:2765–78.
[6] Lee KJ, Inoue M, Otani T, Iwasaki M, Sasazuki S, Tsugane S. Physical activity and risk of colorectal cancer in Japanese men and women: the Japan Public Health Center-based prospective study. Cancer Causes Contr 2007;18:199–209.
[7] Isomura K, Kono S, Moore MA, Toyomura K, Nagano J, Mizoue T, et al. Physical activity and colorectal cancer: the Fukuoka colorectal cancer study. Cancer Sci 2006;97:1099–104.
[8] Otani T, Iwasaki M, Yamamoto S, Sobue T, Hanaoka T, Inoue M, et al. Alcohol consumption, smoking, and subsequent risk of colorectal cancer in middle-aged and elderly Japanese men and women: Japan Public Health Center-based prospective study. Cancer Epidemiol Biomarkers Prev 2003;12:1492–500.
[9] Shimizu N, Nagata C, Shimizu H, Kametani M, Takeyama N, Ohnuma T, et al. Height, weight, and alcohol consumption in relation to the risk of colorectal cancer in Japan: a prospective study. Br J Cancer 2003;88:1038–43.
[10] Otani T, Iwasaki M, Inoue M. Body mass index, body height, and subsequent risk of colorectal cancer in middle-aged and elderly Japanese men and women: Japan public health center-based prospective study. Cancer Causes Contr 2005;16:839–50.
[11] Tsubono Y, Otani T, Kobayashi M, Yamamoto S, Sobue T, Tsugane S. No association between fruit or vegetable consumption and the risk of colorectal cancer in Japan. Br J Cancer 2005;92:1782–4.
[12] Otani T, Iwasaki M, Ishihara J, Sasazuki S, Inoue M, Tsugane S. Dietary fiber intake and subsequent risk of colorectal cancer: the Japan Public Health Center-based prospective study. Int J Cancer 2006;119:1475–80.
[13] Mizoue T, Inoue M, Wakai K, Nagata C, Shimazu T, Tsuji I, et al. Alcohol drinking and colorectal cancer in Japanese: a pooled analysis of results from five cohort studies. Am J Epidemiol 2008;167:1397–406.

[14] Mizoue T, Inoue M, Tanaka K, Tsuji I, Wakai K, Nagata C, et al. Tobacco smoking and colorectal cancer risk: an evaluation based on a systematic review of epidemiologic evidence among the Japanese population. Jpn J Clin Oncol 2006;36:25–39.
[15] Akhter M, Inoue M, Kurahashi N, Iwasaki M, Sasazuki S, Tsugane S. Reproductive factors, exogenous female hormone use and colorectal cancer risk: the Japan Public Health Center-based Prospective Study. Eur J Cancer Prev 2008;17:515–24.
[16] Tamakoshi K, Wakai K, Kojima M, Watanabe Y, Hayakawa N, Toyoshima H, et al. A prospective study of reproductive and menstrual factors and colon cancer risk in Japanese women: findings from the JACC study. Cancer Sci 2004;95:602–7.
[17] Tsugane S. What we know about associations between diet and cancer. JMAJ 2008;51:7.
[18] Freedman AN, Slattery ML, Ballard-Barbash R, Willis G, Cann BJ, Pee D, et al. Colorectal cancer risk prediction tool for white men and women without known susceptibility. J Clin Oncol 2008.
[19] Parkin DM, Olsen AH, Sasieni P. The potential for prevention of colorectal cancer in the UK. Eur J Cancer Prev 2009;18:179–90.
[20] Colditz GA, Atwood KA, Emmons K, Monson RR, Willett WC, Trichopoulos D, et al. Harvard report on cancer prevention volume 4: Harvard cancer risk index. Risk Index Working Group, Harvard Center for Cancer Prevention. Cancer Causes Contr 2000;11:477–88.
[21] Selvachandran SN, Hodder RJ, Ballal MS, Jones P, Cade D. Prediction of colorectal cancer by a patient consultation questionnaire and scoring system: a prospective study. Lancet 2002;360:278–83.
[22] Imperiale TF, Wagner DR, Lin CY, Larkin GN, Rogge JD, Ransohoff DF. Using risk for advanced proximal colonic neoplasia to tailor endoscopic screening for colorectal cancer. Ann Intern Med 2003;139:959–65.
[23] Tsugane S, Sobue T. Baseline survey of JPHC study—design and participation rate. Japan public health center-based prospective study on cancer and cardiovascular diseases. J Epidemiol 2001;11:S24–9.
[24] Iwasaki M, Otani T, Yamamoto S, Inoue M, Hanaoka T, Sobue T, et al. Background characteristics of basic health examination participants: the JPHC study baseline survey. J Epidemiol 2003;13:216–25.
[25] Sasaki S, Kobayashi M, Ishihara J, Tsugane S. Self-administered food frequency questionnaire used in the 5-year follow-up survey of the JPHC study: questionnaire structure, computation algorithms, and area-based mean intake. J Epidemiol/Jpn Epidemiol Assoc 2003;13:S13–22.
[26] Inoue M, Iwasaki M, Otani T, Sasazuki S, Noda M, Tsugane S. Diabetes mellitus and the risk of cancer: results from a large-scale population-based cohort study in Japan. Arch Intern Med 2006;166:1871–7.
[27] Inoue M, Iso H, Yamamoto S, Kurahashi N, Iwasaki M, Sasazuki S, et al. Daily total physical activity level and premature death in men and women: results from a large-scale population-based cohort study in Japan (JPHC study). Ann Epidemiol 2008;18:522–30.
[28] Technology. CfSa, Ministry of Education C., Sports, Science and Technology the Government of Japan, ed. Standard tables of food composition in Japan, the fifth revised and enlarged edition. Tokyo: Printing Bureau, Ministry of Finance, 2005.
[29] Harrell Jr FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361–87.
[30] Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. Int J Epidemiol 1999;28:964–74.
[31] Sauerbrei W, Meier-Hirmer C, Benner A, Royston P. Multivariable regression model building by using fractional polynomials: description of SAS, STATA and R programs. Comput Stat Data Anal 2006;50:3464–85.
[32] Liu J, Hong Y, D'Agostino Sr RB, Wu Z, Wang W, Sun J, et al. Predictive value for the Chinese population of the Framingham CHD risk assessment tool compared with the Chinese Multi-Provincial Cohort Study. JAMA 2004;291:2591–9.
[33] D'Agostino Sr RB, Grundy S, Sullivan LM, Wilson P. Validation of the Framingham coronary heart disease prediction scores: results of a multiple ethnic groups investigation. JAMA 2001;286:180–7.
[34] Therneau TM, Grambsch GP. Expected survival. Modeling survival data: extending the Cox model. Springer; 2004. p. 280.
[35] D'Agostino RB, Nam BH. Evaluation of the performance of survival analysis models: discrimination and calibration measures. In: Balakrishnan NRC, ed. Handbook of statistics, vol. 23. London, England: Elsevier, 2004.
[36] Park Y, Freedman AN, Gail MH, Pee D, Hollenbeck A, Schatzkin A, et al. Validation of a colorectal cancer risk prediction model among white patients age 50 years and older. J Clin Oncol 2009;27:694–8.
[37] Lemeshow S, Hosmer Jr DW. A review of goodness of fit statistics for use in the development of logistic regression models. Am J Epidemiol 1982;115:92–106.
[38] Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. Circulation 1998;97:1837–47.
[39] Wu Y, Liu X, Li X, Li Y, Zhao L, Chen Z, et al. Estimation of 10-year risk of fatal and nonfatal ischemic cardiovascular diseases in Chinese adults. Circulation 2006;114:2217–25.
[40] Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. Springer; 2009.
[41] Kannel WB, McGee DL. Composite scoring—methods and predictive validity: insights from the Framingham Study. Health Serv Res 1987;22:499–535.
[42] Giovannucci E, Rimm EB, Stampfer MJ, Colditz GA, Ascherio A, Willett WC. Intake of fat, meat, and fiber in relation to risk of colon cancer in men. Cancer Res 1994;54:2390–7.
[43] Cross AJ, Leitzmann MF, Gail MH, Hollenbeck AR, Schatzkin A, Sinha R. A prospective study of red and processed meat intake in relation to cancer risk. PLoS Med 2007;4:e325.
[44] McCullough ML, Feskanich D, Stampfer MJ, Giovannucci EL, Rimm EB, Hu FB, et al. Diet quality and major chronic disease risk in men and women: moving toward improved dietary guidance. Am J Clin Nutr 2002;76:1261–71.
[45] Kim MK, Sasaki S, Sasazuki S, Tsugane S. Prospective study of three major dietary patterns and risk of gastric cancer in Japan. Int J Cancer 2004;110:435–42.
[46] Lee KJ, Inoue M, Otani T, Iwasaki M, Sasazuki S, Tsugane S. Colorectal cancer screening using fecal occult blood test and subsequent risk of colorectal cancer: a prospective cohort study in Japan. Cancer Detect Prev 2007;31:3–11.
[47] Chen J, Pee D, Ayyagari R, Graubard B, Schairer C, Byrne C, et al. Projecting absolute invasive breast cancer risk in white women with a model that includes mammographic density. J Natl Cancer Inst 2006;98:1215–26.
[48] Barlow WE, White E, Ballard-Barbash R, Vacek PM, Titus-Ernstoff L, Carney PA, et al. Prospective breast cancer risk prediction model for women undergoing screening mammography. J Natl Cancer Inst 2006;98:1204–14.
[49] Rosner BA, Colditz GA, Webb PM, Hankinson SE. Mathematical models of ovarian cancer incidence. Epidemiology 2005;16:508–15.
[50] Cronin KA, Gail MH, Zou Z, Bach PB, Virtamo J, Albanes D. Validation of a model of lung cancer risk prediction among smokers. J Natl Cancer Inst 2006;98:637–40.