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