ORIGINAL ARTICLE

Ethnicity and Dietary Practices as Colorectal Cancer Risk Predictors: A Retrospective Case-control Study in Sabah, Malaysia

Wen-Li Tee[1,2], *Fredie Robinson[1], Richard Avoi[1], Prabakaran Dhanaraj[3], Nirmal Kaur[4], Nur Hasanah Sanudin[4]

[1] Department of Public Health Medicine, Faculty of Medicine and Health Sciences, Universiti Malaysia Sabah, 88400 Kota Kinabalu, Sabah, Malaysia
[2] Primary Health Sector, Public Health Division, Sabah State Health Department, 88300 Kota Kinabalu, Sabah, Ministry of Health, Malaysia
[3] Kota Kinabalu Area Health Office, Sabah State Health Department, 88300 Kota Kinabalu, Sabah, Ministry of Health, Malaysia
[4] Non-Communicable Diseases Sector, Public Health Division, Sabah State Health Department, 88300 Kota Kinabalu, Sabah, Ministry of Health, Malaysia

ABSTRACT

Introduction: The association between colorectal cancer (CRC), ethnicity, and dietary practices have been well studied. However, limited studies have been conducted to assess dietary practices and ethnicity in Sabah on risk of CRC. This study aimed to assess the risk and protective factors in dietary practices and the inclusion of ethnicity and dietary practices as risk predictors for CRC. **Methods:** 148 CRC patients, 609 controls were recruited in this case-control study. Logistic regression analyses were performed to determine significant predictors of CRC. Prediction model was computed using Logistic Regression (LR) and C5 Decision Tree algorithms and compared. **Results:** Age 60-69 (aOR = 7.44, 95% CI = 3.69-15.00); male (aOR = 4.49, 95% CI = 2.67-7.54), Chinese (aOR = 32.32, 95% CI = 7.20-145.13); moderate physical activity (aOR = 3.67, 95% CI = 2.03-6.63), pickled mango (aOR = 5.66, 95% CI = 1.62-19.81), pork (aOR = 2.29, 95% CI = 1.09-4.79) increased the odds of developing CRC. No comorbidities (aOR = 0.53, 95% CI = 0.31-0.91), tertiary education attainment (aOR = 0.18, 95% CI = 0.07-0.43) were protective against CRC. Hosmer-Lemeshow test indicated good fit of the model (p =.946) and excellent discriminatory power (AUC=0.877). LR prediction model demonstrated better overall accuracy (89.2%), discriminatory power (AUC=0.82), sensitivity (77%), and specificity (91%) than the C5 Model. **Conclusion:** Frequent consumption of pickled mangoes and pork increased CRC risk among the Sabah population. Inclusion of ethnicity and dietary practices as predictors could potentially improve risk stratification of the Sabah population for early CRC screening.

Keywords: Colorectal cancer; Ethnicity; Dietary practices; Prediction; Screening

Corresponding Author:
Fredie Robinson, PhD
Email: freddie@ums.edu.my
Tel: +6088320000

INTRODUCTION

Colorectal cancer (CRC) is a prevalent global disease with a 60% increase in burden, reaching 2.2 million new cases and 1.1 million deaths by 2030 [1]. In Sabah, CRC is the second most common type of cancer, with Chinese having the highest incidence, followed by Bajau among males and Murut among females [1]. Most CRC in Sabah remain diagnosed at late stage III or IV, resulting in an unfavorable prognosis, lower survival rates and requires higher treatment costs, even with the availability of guidelines and policies [1–3].

Studies have demonstrated that the incidence of CRC varies among ethnicities [1, 2, 4–6]. Ethnicity influences lifestyle and dietary practices, and Sabah's diverse culture and dietary practices vary across ethnic groups, which may play a role in addressing CRC susceptibility [5, 7, 8]. There are 33 officially recognized indigenous ethnic groups in Sabah: Kadazan-Dusun forms the largest indigenous ethnic group, combining two indigenous tribes at 23.6%, followed by Bajau, Murut, and other smaller indigenous ethnic groups. The Chinese form the largest non-indigenous group in Sabah [8].

Diet is a known contributing factor to CRC. This study considers it crucial to determine the association between dietary practices common in Sabah and the development of CRC, as there are limited data on this research problem. Food preparation methods such as fermentation and pickling are commonly practiced and consumed in Sabah. Preserved food in Sabah includes jeruk tuhau (pickled wild ginger), jeruk bambangan (pickled wild mango), ikan masin (dried salted fish), ikan liking (wet salted fish), bosou ikan (fermented fish), bosou sayur (fermented vegetables), etc [9, 10]. According to the International Agency for Research and Cancer Group (IARC), preserved foods, such as processed meat and salted fish, are classified as carcinogenic to humans [11, 12]. High-temperature cooking methods, such as frying, grilling, and smoking, are also carcinogenic to the human digestive system [12–14].

CRC is preventable and treatable if detected at an early stage. Malaysia's Ministry of Health aims to control the CRC burden by early detection and downstaging CRC at the time of diagnosis. However, the current policy in Malaysia only targets those aged 50 years and above and those with a family history of CRC. Ethnicity and dietary practices were excluded as screening criteria for CRC risk stratification [2]. The inclusion of ethnicity and dietary practices as predictors in risk stratification; therefore, could potentially improve early detection and diagnosis, by capturing a larger and more specific population at risk for CRC. Furthermore, the majority of cancer risk prediction scores were developed for the Western population, as the majority of cancer research was conducted in Western countries. Such risk prediction scores may not accurately predict cancer risks for local ethnic groups.

To enhance the early detection and diagnosis of CRC, this study aimed to explore the risk and protective factors in dietary practices associated with CRC that are specific to Sabah, particularly investigating the role of traditional preserved and fermented food. The study also aimed to investigate the role of ethnicity on the risk of CRC development among the Sabah population. Following this, the study aimed to develop a prediction model by incorporating ethnicity and dietary practices as CRC risk predictors.

MATERIALS AND METHODS

An unmatched, retrospective observational case-control study involving 148 CRC patients and 609 cancer-free participants aged 40–75 years was conducted to explore the risk and protective factors of dietary practices on CRC and assess ethnicity and dietary practices as CRC risk predictors in Sabah, Malaysia. Patients histopathologically diagnosed with CRC between 2018 and 2022 were recruited from the Colorectal Unit Hospital Queen Elizabeth (HQE), Sabah. Cancer-free participants in the control group were defined as individuals without prior CRC or polyp diagnosis; asymptomatic individuals tested negative on the fecal occult blood test (FOBT) and normal colonoscopy findings. We stratified the controls by recruiting them from primary health clinics in different districts to ensure that individuals of different ethnicities were recruited. The intent of involving different districts was to ensure that this study recruits different ethnic groups as dietary practices vary based on the geography and demographic distribution in Sabah. Furthermore, controls were not recruited from the HQE owing to visiting restrictions to the hospital during the COVID-19 pandemic. The primary health clinics involved in the recruitment of controls were located in different districts of Sabah, namely Kota Kinabalu, Sipitang, Penampang, Kudat, Kota Marudu, Tenom, Ranau, Beluran, Lahad Datu, Semporna and Tawau. The ethnic groups recruited in this study were Kadazan, Dusun, Bajau, Murut, Rungus, Sungai, Brunei, Bugis, Lundayeh, Suluk, Bisaya, Kedayan, Tidung, Jawa, Iranun, Cagayan, Banjar, Chinese and Malay. Participants with a family history of CRC, familial adenomatous polyposis, Lynch Syndrome, Peutz-Jegher Syndrome, Juvenile Polyposis, MUTYH-associated polyposis, special diet, or diagnosed with another form of malignancy were excluded from this study.

An interviewer-guided questionnaire was developed and validated for this study, consisting of 4 sections including Section 1 Socio-demographic Characteristics; Section 2 Lifestyle, on physical activity, smoking, alcohol consumption; Section 3 Dietary Practices; Section 4 Anthropometric Measurements. Section 2 consists of 14 types of alcoholic beverages and 6 types of tobacco products. Section 3 consists of 137 food items and common local cooking methods. This questionnaire was then used to collect data from September 2022 to October 2022. Participants were asked by trained interviewers using section 3 of the questionnaire to provide the frequency of consumption and the total number of servings consumed in each meal based on the portion size for each food item, with a 45-minute response time. Frequency of consumption for each of the food items were categorised into "never or less than once a month/ once a month/ 2-3 times a month/ once a week/ 2-4 times a week/ 5-6 times a week/ once a day/ 2-3 times a day/ 4-5 times a day". The food portion size standardized with a meal size graphical reference using common Malaysian household utensils was based on the National Health and Morbidity Survey (NHMS) 2019 questionnaire [15]. Section 4 consists of questions where height and weight was measured to obtain body mass index (BMI). Participants were required to answer the questionnaire by recalling the diet they have consumed for the past one year. Paticipants in the case groups were requested to recall their habitual diet one year prior to being diagnosed with CRC. Height, weight and BMI was obtained by measuring the control groups, while the measurements for CRC patients were obtained from their hospital clerking sheets when they were first diagnosd with CRC.

Descriptive analyses were performed to analyse variables such as demography, socioeconomic characteristics, and anthropometric parameters. Proportions were computed for categorical variables. Chi-square test of association between each baseline characteristics and the outcomes were computed. Data from the food intake frequency were re-coded into two levels - < 2 times/week; ≥ 2 times/week for the computation on logistic regression. Logistic regression analyses were performed to estimate the exposure odds ratios (ORs) and 95% confidence intervals (CIs). Univariate analysis was performed separately to predict the effects of each independent variable on the likelihood of CRC development. Variables with p ≤0.05 were selected to proceed with multivariable analysis. The sample size to compute logistic regression analyses in this study was based on the rule of event per variable (EPV) to prevent overestimation of the effect measure. A Malaysian study recommended EPV of 50 with the formula $n = 100 + 50i$, where i refers to the number of independent variables in the final model, a minimum sample size of 500 is required to perform a logistic regression analysis [16]. Based on this formula, our study has selected 12 significant risk predictors to proceed to computing multivariable analysis based on the sample size acquired in this study. Multivariable analysis was then performed using backward and forward procedures to select independent risk factors that best predicted CRC. Interaction terms and multicollinearity were assessed after obtaining a preliminary main-effect model. The overall model fitness, performance, and discriminatory power were examined for the final model using the Hosmer-Lemeshow test and the Area under the Receiver Operating Characteristic curve (AUC).

CRC risk prediction models were computed using two predictive modelling algorithms and performance between two models were compared to see which model is best suited for CRC risk prediction. Both Logistic Regression (LR) and C5 Decision Tree are classification techniques commonly used in predictive modelling, with LR being the conventional statistical method commonly used in the medical field, and the decision tree based technique of the C5 is a non-parametric supervised machine learning algorithm used for both classification and regression tasks. The input variables were selected based on the significant predictors obtained from the multivariable analysis. The dataset for all prediction models was partitioned into 70% for the training dataset and 30% for the testing dataset. The intent to split the original data into smaller datasets was to explore the characteristics of the data, 'train' and create the model using the training dataset, and then measure the overall model performance using the testing dataset. The overall accuracy, sensitivity, specificity, and area under the Receiver Operating Characteristic curve (AUC) of the models were assessed and compared. All statistical analyses, including the LR model, were computed using IBM-SPSS software version 27. The C5 Decision Tree Model was computed using IBM-SPSS Modeler version 18.3.

Informed consent was obtained from all research participants for this study. Ethical approval for this study (NMRR ID-22-01797-G6Y) was provided by the Medical Research and Ethics Committee (MREC) Ministry of Health, Malaysia.

RESULTS

Sociodemographic characteristics

Among the participants, 326 (43.1%) were male and 431 (56.9%) were female. The majority of participants in this study were in the 40-49 age group (47.4%). By ethnicity, the majority of participants were Bajau (22.8%). Among the cases, CRC patients were mostly Chinese (23.6%), followed by Dusun (19.6%), Bajau (16.2%), Kadazan (11.5%), and other Sabah minority ethnic groups (17.6%). The majority of participants were government employees (29.8%). Most of the study population falls under the lower household income range of RM<2,500 (55.2%) and RM 2,501-4,849 (33.6%). Most participants attained secondary education (42.6%). More than half of the study population had no comorbidities (51.8%). As for the lifestyle habits, most participants engaged in vigorous physical activity (48.5%), never smoker (79.2%), did not consume alcohol (73.6%), and within normal range of body mass index (BMI) (40.3%). The sociodemographic characteristics of the 757 participants are shown in Table I.

Logistic Regression Analyses

Based on the findings of the univariate analysis, variables Age, Gender, Ethnicity, Education Level, Co-morbidities, Physical Activity, Smoking Status, Alcohol Consumption, Body Mass Index (BMI), and Consumption of Pickled Mango, Pork and Mutton had p-value ≤0.05. Although the findings from the univariate analysis demonstrated more statistically significant variables, these 12 variables were selected based on the rule of event per variable (EPV) to prevent overestimation of the effect measure. The results of the univariate analysis are presented in Table II.

Eight predictors remained statistically significant after the multivariable analysis. The overall model fit for the final model with the Hosmer-Lemeshow test demonstrated a good fit (P =.946). The AUC showed excellent discrimination (AUC = 0.877). Cook's influential statistics showed that there were no outliers. Based on the final model, the odds of developing

Table I : Sociodemographic characteristics (n=757)

Variable	Cases (n=148), n(%)	Controls (n=609), n(%)	P [1]
Age group			<.001
40-49	24 (16.2)	335 (55)	
50-59	41 (27.7)	179 (29.4)	
60-69	65 (43.9)	81 (13.3)	
70-79	18 (12.2)	14 (2.3)	
Gender			<.001
Male	96 (64.9)	230 (37.8)	
Female	52 (35.1)	379 (62.2)	
Ethnicity			<.001
Kadazan	17 (11.5)	47 (7.7)	
Dusun	29 (19.6)	87 (14.3)	
Murut	5 (3.4)	71 (11.7)	
Bajau	24 (16.2)	148 (24.3)	
Brunei	4 (2.7)	17 (2.8)	
Chinese	35 (23.6)	21 (3.4)	
Malay	2 (1.4)	21 (3.4)	
Other Sabah Minority Ethnic Groups	26 (17.6)	113 (18.6)	
Others	2 (1.4)	8 (1.3)	
Rungus	4 (2.7)	76 (12.5)	
Occupation			<.001
government employee	14 (9.5)	212 (34.8)	
private employee	8 (5.4)	54 (8.9)	
self-employed	32 (21.6)	152 (24.9)	
home maker	23 (15.5)	131 (21.5)	
retiree	21 (14.2)	6 (0.9)	
unemployed	50 (33.8)	54 (8.9)	
Household income (RM)			<.063
< 2500	96 (64.9)	322 (52.9)	
2501-4849	40 (27.0)	214 (35.1)	
4850-10959	9 (6.1)	60 (9.9)	
≥ 10960	3 (2.0)	13 (2.1)	
Education level			<.001
primary	28 (18.9)	73 (12.0)	
secondary	68 (45.9)	255 (41.9)	
tertiary	25 (16.9)	237 (38.9)	
none	27 (18.2)	44 (7.2)	
Co-morbidities			<.001
no	35 (23.6)	357 (58.6)	
yes	113 (76.4)	252 (41.4)	

Variable	Cases (n=148), n(%)	Controls (n=609), n(%)	P [1]
Physical activity			<.001
sedentary	13 (8.8)	64 (10.5)	
low	27 (18.2)	61 (10.0)	
moderate	64 (43.2)	161 (26.4)	
vigorous	44 (29.7)	323 (53.0)	
Smoking status			<.001
light smoker	38 (25.7)	103 (16.9)	
moderate smoker	8 (5.4)	5 (0.8)	
heavy smoker	1 (0.7)	2 (0.3)	
never smoker	101 (68.2)	499 (81.9)	
Alcohol consumption			<.001
yes	58 (39.2)	142 (23.3)	
no	90 (60.8)	467 (76.7)	
BMI			.001
underweight	13 (8.8)	13 (2.1)	
normal	55 (37.2)	250 (41.1)	
overweight	54 (36.5)	234 (38.4)	
obesity	26 (17.6)	112 (18.4)	

[1] Chi-square test of association between each characteristics and the outcomes.

Table II : Unadjusted predictors of colorectal cancer

Variable		Case (n= 148)	Control (n=609)	Crude OR	(95% CI OR)	χ^2 stat. (df)[a]	P value[a]
Age						115.77 (3)	<.001
	40-49	24	335	1			
	50-59	41	179	3.2	(1.87 ; 5.46)	18.10 (1)[b]	<.001[b]
	60-69	65	81	11.2	(6.61 ; 18.98)	80.64 (1)[b]	<.001[b]
	70-79	18	14	17.95	(7.97 ; 40.42)	48.57 (1)[b]	<.001[b]
Gender							
	Male	96	230	3.04	(2.09 ; 4.43)	35.49 (1)	<.001
	Female	52	379	1			
Ethnicity						83.76 (9)	<.001
	Kadazan	17	47	6.87	(2.18 ; 21.67)	10.82 (1)[b]	.001[b]
	Dusun	29	87	6.33	(2.13 ; 18.83)	11.02 (1)[b]	.001[b]
	Murut	5	71	1.34	(0.35 ; 5.18)	0.18 (1)[b]	.673[b]
	Bajau	24	148	3.08	(1.03 ; 9.20)	4.06 (1)[b]	.044[b]
	Brunei	4	17	4.47	(1.02 ; 19.68)	3.92 (1)[b]	.048[b]
	Chinese	35	21	31.67	(10.11 ; 99.19)	35.18 (1)[b]	<.001[b]
	Malay	2	21	1.81	(0.31 ; 10.57)	0.43 (1)[b]	.51[b]
	Other Sabah Minority Ethnic Groups	26	113	4.37	(1.47 ; 13.03)	7.01 (1)[b]	.008[b]
	Others	2	8	4.75	(0.75 ; 30.12)	2.73 (1)[b]	.098[b]
	Rungus	4	76	1			

Variable		Case (n= 148)	Control (n=609)	Crude OR	(95% CI OR)	χ^2 stat. (df)[a]	P value[a]
Education level						37.05 (3)	<.001
	primary	28	73	0.63	(0.33 ; 1.19)	2.02 (1)[b]	.155[b]
	secondary	68	255	0.44	(0.25 ; 0.75)	8.86 (1)[b]	.003[b]
	tertiary	25	237	0.17	(0.09 ; 0.32)	29.82 (1)[b]	<.001[b]
	none	27	44	1			
Co-morbidities							
	no	35	357	0.22	(0.15 ; 0.33)	60.49 (1)	<.001
	yes	113	252	1			
Physical activity						31.79 (3)	<.001
	sedentary	13	64	1.49	(0.76 ; 2.93)	1.35 (1)[b]	.246[b]
	low	27	61	3.25	(1.87 ; 5.64)	17.52 (1)[b]	<.001[b]
	moderate	64	161	2.92	(1.90 ; 4.48)	24.07 (1)[b]	<.001[b]
	vigorous	44	323	1			
Smoking status						18.72 (3)	<.001
	light smoker	38	103	1.82	(1.19 ; 2.80)	7.52 (1)[b]	.006[b]
	moderate smoker	8	5	7.91	(2.53 ; 24.66)	12.69 (1)[b]	<.001[b]
	heavy smoker	1	2	2.47	(0.22 ; 27.50)	0.54 (1)[b]	.462[b]
	never smoker	101	499	1			
Alcohol consumption							
	yes	58	142	2.06	(1.41 ; 3.01)	13.38 (1)	<.001
	no	90	467	1			
BMI						13.60 (3)	.004
	underweight	13	13	4.31	(1.79 ; 10.38)	10.60 (1)[b]	.001[b]
	normal	55	250	0.95	(0.57 ; 1.59)	0.04 (1)[b]	.839[b]
	overweight	54	234	0.99	(0.59 ; 1.67)	0.001 (1)[b]	.982[b]
	obesity	26	112	1			
Preserved food							
bosou sayur (fermented vegetable)	≥ 2 times/week	6	31	0.79	(0.32 ; 1.93)	0.29 (1)	.592
	< 2 times/week	142	578	1			
bosou ikan (fermented fish)	≥ 2 times/week	8	29	1.14	(0.51 ; 2.55)	0.10 (1)	.748
	< 2 times/week	140	580	1			
pickled mango	≥ 2 times/week	9	13	2.97	(1.24 ; 7.08)	5.46(1)	.019
	< 2 times/week	139	596	1			

Variable		Case (n= 148)	Control (n=609)	Crude OR	(95% CI OR)	χ^2 stat. (df)[a]	P value[a]
dried salted fish	≥ 2 times/week	18	77	0.96	(0.55 ; 1.65)	0.03 (1)	.874
	< 2 times/week	130	532	1			
tuhau (pickled wildginger)	≥ 2 times/week	12	27	1.9	(0.94 ; 3.85)	2.95(1)	.086
	< 2 times/week	136	582	1			
jeruk bambangan (pickled wild mango)	≥ 2 times/week	3	12	1.03	(0.29 ; 3.70)	0.002 (1)	.965
	< 2 times/week	145	597	1			
salted egg	≥ 2 times/week	10	33	1.27	(0.61 ; 2.63)	0.38 (1)	.537
	< 2 times/week	138	576	1			
ikan liking (wet salted fish)	≥ 2 times/week	7	19	1.54	(0.64 ; 3.74)	0.86 (1)	.354
	< 2 times/week	141	590	1			
salted vegetable	≥ 2 times/week	4	10	1.66	(0.52 ; 5.38)	0.67 (1)	.413
	< 2 times/week	144	599	1			
Meat							
pork	≥ 2 times/week	59	41	3.57	(2.28 ; 5.60)	28.79 (1)	<.001
	< 2 times/week	550	107	1			
wild boar	≥ 2 times/week	5	8	2.63	(0.85 ; 8.15)	2.53 (1)	.112
	< 2 times/week	143	601	1			
mutton	≥ 2 times/week	5	6	3.51	(1.09 ; 11.68)	3.84 (1)	.05
	< 2 times/week	143	603	1			
beef	≥ 2 times/week	6	47	0.51	(0.21 ; 1.21)	2.77 (1)	.096
	< 2 times/week	142	562	1			
processed meat	≥ 2 times/week	21	80	1.09	(0.65 ; 1.84)	0.113 (1)	.737
	< 2 times/week	127	529	1			
Vegetables							
ulam raja	≥ 2 times/week	10	82	0.47	(0.24 ; 0.92)	5.67 (1)	.017
	< 2 times/week	138	527	1			

Variable		Case (n= 148)	Control (n=609)	Crude OR	(95% CI OR)	χ^2 stat. (df)[a]	P value[a]
sayur pakis (fern)	≥ 2 times/week	32	182	0.65	(0.42 ; 0.99)	4.18 (1)	.041
	< 2 times/week	116	427	1			
cassava leaves	≥ 2 times/week	32	147	0.87	(0.56 ; 1.34)	0.424 (1)	.515
	< 2 times/week	116	462	1			
losun	≥ 2 times/week	13	50	1.08	(0.57 ; 2.04)	0.051 (1)	.822
	< 2 times/week	135	559	1			
Tubers							
potato	≥ 2 times/week	34	98	1.56	(1.00 ; 2.41)	3.706 (1)	.054
	< 2 times/week	114	511	1			
cassava	≥ 2 times/week	16	71	0.92	(0.52 ; 1.63)	0.085 (1)	.777
	< 2 times/week	132	538	1			
Soy, legumes							
taufufa (soy beancurd)	≥ 2 times/week	5	33	0.6	(0.23 ; 1.59)	1.139 (1)	.286
	< 2 times/week	143	576	1			
tempe (fermented soybean)	≥ 2 times/week	8	54	0.59	(0.27 ; 1.26)	2.08 (1)	.149
	< 2 times/week	140	555	1			
tofu	≥ 2 times/week	37	83	2.11	(1.36 ; 3.27)	10.51 (1)	.001
	< 2 times/week	111	526	1			
legume	≥ 2 times/week	5	54	0.36	(0.14 ; 0.92)	5.96 (1)	.015
	< 2 times/week	143	555	1			
Dairy products							
yogurt	≥ 2 times/week	1	26	0.2	(0.02 ; 1.13)	6.18 (1)	.013
	< 2 times/week	147	583	1			
milk	≥ 2 times/week	50	187	1.15	(0.79 ; 1.69)	0.519 (1)	.471
	< 2 times/week	98	422	1			
Local delicacy							
sago grub	≥ 2 times/week	3	5	2.5	(0.59 ; 10.58)	1.401 (1)	.237
	< 2 times/week	145	604	1			
Food preparation methods							
slow cooking	≥ 2 times/week	18	126	0.53	(0.31 ; 0.90)	6.116 (1)	.013
	< 2 times/week	130	483	1			

Variable		Case (n= 148)	Control (n=609)	Crude OR	(95% CI OR)	χ^2 stat. (df)[a]	P value[a]
smoke	≥ 2 times/week	2	24	0.33	(0.08 ; 1.43)	2.95 (1)	.086
	< 2 times/week	146	585	1			
boil	≥ 2 times/week	100	465	0.65	(0.44 ; 0.95)	4.67 (1)	.031
	< 2 times/week	48	144	1			
deep fry	≥ 2 times/week	77	428	0.46	(0.32 ; 0.66)	17.16 (1)	<.001
	< 2 times/week	71	181	1			
acid cooking	≥ 2 times/week	30	4	0.54	(0.19 ; 1.55)	1.54 (1)	.214
	< 2 times/week	144	579	1			
grill/bake	≥ 2 times/week	11	80	0.53	(0.28 ; 1.03)	4.05(1)	.044
	< 2 times/week	137	529	1			
stew	≥ 2 times/week	31	150	0.81	(0.52 ; 1.26)	0.91(1)	.340
	< 2 times/week	117	459	1			
raw	≥ 2 times/week	5	22	0.93	(0.35 ; 2.51)	0.02(1)	.890
	< 2 times/week	143	587	1			
stir fry	≥ 2 times/week	123	531	0.72	(0.44 ; 1.18)	1.61(1)	.204
	< 2 times/week	25	78	1			
steam	≥ 2 times/week	47	223	0.81	(0.55 ; 1.18)	1.24(1)	.265
	< 2 times/week	101	386	1			

Unadj. OR = Unadjusted Odds Ratio [a] Likelihood Ratio (LR) test, [b] Wald test.

Table III : Adjusted predictors of colorectal cancer

Variable		Adj. OR	(95% CI OR)	χ^2 stat. (df)[a]	P value[a]
Age				35.33 (3)	<.001
	40-49	1.00			
	50-59	3.41	(1.76 ; 6.60)	13.29 (1)[b]	<.001[b]
	60-69	7.44	(3.69 ; 15.00)	31.48 (1)[b]	<.001[b]
	70-79	6.38	(2.01 ; 20.20)	9.92 (1)[b]	.002[b]
Gender					
	Male	4.49	(2.67 ; 7.54)	35.44 (1)	<.001
	Female	1.00			
Ethnicity				43.59 (9)	<.001
	Kadazan	13.84	(3.25 ; 58.96)	12.62 (1)[b]	.001[b]
	Dusun	10.51	(2.81 ; 39.39)	12.18 (1)[b]	<.001[b]

Variable		Adj. OR	(95% CI OR)	χ^2 stat. (df)[a]	P value[a]
	Murut	1.70	(0.35 ; 8.29)	0.43 (1)[b]	.511[b]
	Bajau	6.50	(1.75 ; 24.08)	7.84 (1)[b]	.005[b]
	Brunei	4.26	(1.02 ; 19.68)	2.48 (1)[b]	.116[b]
	Chinese	32.32	(7.20 ; 145.13)	20.57 (1)[b]	<.001[b]
	Malay	5.40	(0.72 ; 40.60)	2.69 (1)[b]	.101[b]
	Other Sabah Minority Ethnic Groups	6.61	(1.78 ; 24.55)	7.96 (1)[b]	.005[b]
	Others	7.16	(0.62 ; 82.85)	2.49 (1)[b]	.115[b]
	Rungus	1.00			
Education level				21.26 (3)	<.001
	primary	0.61	(0.25 ; 1.47)	1.21 (1)[b]	.271[b]
	secondary	0.54	(0.25 ; 1.18)	2.37 (1)[b]	.124[b]
	tertiary	0.18	(0.07 ; 0.43)	14.70 (1)[b]	<.001[b]
	none	1.00			
Co-morbidities					
	no	0.53	(0.31 ; 0.91)	5.39 (1)	.020
	yes	1.00			
Physical activity				23.06 (3)	<.001
	sedentary	1.02	(0.41 ; 2.53)	0.00 (1)[b]	.963[b]
	low	1.20	(0.55 ; 2.59)	0.20 (1)[b]	.652[b]
	moderate	3.67	(2.03 ; 6.63)	18.56 (1)[b]	<.001[b]
	vigorous	1.00			
Preserved food					
pickled mango	≥ 2 times/week	5.66	(1.62 ; 19.81)	6.74 (1)	.009
	< 2 times/week	1.00			
Meat					
pork	≥ 2 times/week	2.29	(1.09 ; 4.79)	4.72 (1)	.030
	< 2 times/week	1.00			

Adj. OR = adjusted odds ratio, [a] likelihood ratio (LR) test, [b] Wald test.

CRC increased with age; those aged 60-69 years had 7.4 times the odds of developing CRC (aOR =7.44, 95% CI = 3.69 - 15.00). Males had 4.5 times the odds of developing CRC (aOR =4.49, 95% CI = 2.67 - 7.54). Being Chinese increased the odds of developing CRC 32 times (aOR =32.32, 95% CI = 7.20 - 145.13). Among the Sabah ethnic groups, Kadazan had 13.8 times the odds of developing CRC (aOR =13.84, 95% CI = 3.25 - 58.96), 10 times more likely among Dusun (aOR =10.51, 95% CI = 2.81 - 39.39) and 6 times more likely among Bajau (aOR =6.50, 95% CI = 1.75 - 24.08). Those with higher educational attainment at the tertiary level were 82% less likely to develop CRC (aOR =0.18, 95% CI = 0.07 - 0.43). There was a 47% decrease in CRC incidence among those without comorbidities (aOR =0.53, 95% CI = 0.31 - 0.91). Frequent consumption of pickled mango more than 2 times a week had five times the odds of developing CRC (aOR =5.66, 95% CI = 1.62 - 19.81), and consumption of pork had 2 times likely to develop CRC (aOR =2.29, 95% CI = 1.09 - 4.79). The results of the final model are presented in Table III.

Colorectal cancer risk predictive modelling

The significant predictors obtained from the multivariable analysis included age, gender, ethnicity,

Table IV : Results for the LR model

Performance measures	Models	
	Training (n=535)	Testing (n=222)
Discrimination (AUC)	0.869 (0.84-0.90)	0.818 (0.78-0.86)
Calibration (H-L test)	χ^2 = 5.97 (P =.651)	χ^2 = 3.84 (P =.872)
Nagelkerke R square (R^2)	0.531	0.539
Overall accuracy (%)	87.9	89.2
Sensitivity	0.75	0.77
Specificity	0.90	0.91

Table V : Results for the C5 model

Performance Measures	Models	
	Training (n=533)	Testing (n=224)
Overall Accuracy	84.43%	82.59%
AUC	0.78	0.73
Sensitivity	0.68	0.63
Specificity	0.86	0.84
Predictors included in the model (predictor importance)	Age (0.76) Frequent pork intake (0.23) Pickled mango (0.01)	

Table VI : Comparison from training and testing datasets for the prediction models

Models	Training		Testing	
	Logistic Regression	C5	Logistic Regression	C5
Overall Accuracy (%)	87.9	84.4	89.2	82.6
AUC	0.87	0.78	0.82	0.73
Sensitivity (%)	75	68	77	63
Specificity (%)	90	86	91	84
No. of predictors included	8	3	8	3

education level, comorbidities, physical activity, and frequent intake of pork and pickled mangoes, were selected as the input variables for the predictive modelling. Both the LR and C5 models were computed and compared. The LR model demonstrated an overall accuracy of 89.2%, with excellent discrimination (AUC = 0.818). The sensitivity and specificity of this model were 77% and 91%, respectively. All the eight predictors were included in the model. Tables IV and V show the results of the training and testing datasets for the LR and C5 models.

The C5 model demonstrated an overall accuracy of 82.6%, with acceptable discrimination (AUC = 0.73). The sensitivity and specificity of this model are 63% and 84%, respectively. Three predictors were included in this model, of which age was the most important predictor.

By comparing both prediction models, we concluded that the LR model was better at predicting the risk of CRC. Table VI shows a comparison of the training and testing datasets for both the prediction models.

DISCUSSION

This study revealed that frequent consumption of pickled mangoes and pork is associated with increased risk of developing CRC. Pickled mangoes were the only food item from the preserved food group that had a significant effect on CRC in this study. This agrees with the findings from IARC where pickled food is known to be carcinogenic due to the N-nitroso compounds found in preserved food [17]. A cohort study conducted in Finland suggested that preserved food such as salted or smoked fish, cured meat contains N-nitroso compounds and consumption of such food increased the risk of CRC [18]. Pickled food also increase the risk of other gastro-intestinal cancer, such as the findings in a meta-analysis where pickled food intake is associated with increased risk of esophageal cancer [19]. Studies from the United States and Germany support the findings that red meat consumption increased the risk of CRC. Pork is considered red meat which is a rich source of heme iron and fatty acids, and frequent consumption promotes carcinogenesis [20, 21].

Several studies have revealed that the consumption of vegetables reduces the risk of developing CRC. Fibres obtained from vegetable intake may enhance gut motility, reducing transit time in the intestinal tract, thus reducing carcinogens contact time from the stool [22, 23]. Our study demonstrated that ulam raja and sayur pakis (ferns) were protective against the risk of CRC. A Malaysian study supported this evidence, where consumption of traditional vegetables found in Borneo such as ulam, tuhau and ferns may prevent oxidative damage and reduce cancer risk [24]. However, limited research were done on these traditional herbs and vegetables to determine the dose-response relationship between traditional vegetable intake and cancer risk.

Several studies have demonstrated that soy products may reduce the risk of CRC [25–27], except one study which suggested there were no evidence of protective effect from soy consumption against CRC [28]. Our study found that frequent consumption of legumes lowered the risk of CRC by 64%. However, frequent consumption of tofu was noted to have a 2-fold increased risk of developing CRC, which was in contrast to the findings from other studies. Possible explanation on other contributing factors that increased the risk of CRC include the process of tofu making, additional ingredients used in making tofu such as preservatives. The contrasting results based on our findings in tofu consumption warrants a further investigation. Several studies have shown that fermented foods and drinks may have anticancer effects [23, 29–32]. However, local foods such as tempe (fermented soybeans), bosou ikan (fermented fish), and bosou sayur (fermented vegetables) demonstrated no significant effect on CRC in this study.

Several studies have demonstrated that food preparation and cooking methods may play an important role in influencing CRC risk. Methods such as pickling, curing, smoking, fermentation, frying, boiling and grilling have different influence towards development of CRC. Cooking methods with high temperature such as frying and grilling meat can lead to the formation of carcinogenic chemicals such as polycyclic aromatic hydrocarbons (PAH) and heterocyclic aromatic amines (HAC) [13, 33–35]. Our study found that slow cooking significantly lowers the risk of CRC by 47%, while boiling lowers the risk of CRC by 35%. These findings suggest that both slow cooking and boiling possibly uses lower temperature to cook dishes, likely reducing the formation of carcinogenic chemicals. However, our study observed an inverse association between deep frying and CRC risk. This finding is in contrast to most studies as cooking methods with higher temperature tends to lead to the formation of carcinogenic chemicals that can increase the risk of CRC. Except one study which observed higher frequency consumption of grilled or barbeque hamburgers contributed to a reduction in CRC risk [20].

Our study revealed that the Chinese were the non-indigenous group with the highest risk of developing CRC, followed by the indigenous ethnic groups Kadazan, Dusun, and Bajau. The incidence for the indigenous ethnic groups was in contrast with the Sabah State Cancer Registry 2012-2016, where Bajau had the highest incidence among the male population and Murut among females [1]. Based on these findings, it is clear that certain ethnic groups have different CRC risk levels. Several studies have considered ethnicity as a determining factor in CRC development [36, 37]. The disproportionate risk of CRC in certain ethnicities demonstrated that ethnicity potentially plays a role in addressing cancer susceptibility as the population within one ethnic group shares common genetic features, and common cultures within the same ethnicity may influence one's lifestyle and dietary practices. This further suggests that specific screening recommendations should be tailored to each of the ethnicities in Sabah as special attention should be given in addressing the high-risk ethnic groups. For instance, high-risk ethnicity should be given earlier CRC screening as compared to low-risk ethnicity. Therefore, the inclusion of ethnicity in risk stratification for early CRC screening is important for the prevention and early detection of CRC.

Our study revealed a 4-fold increased risk of developing CRC among the male population. Similar trend was observed in Malaysia, males are more commonly affected with CRC, accounting for age-standardised rate of 14.8 per 100,000 population [38]. This phenomenon explains the possibility that males are prone to adopt a riskier lifestyle such as alcohol consumption and tobacco smoking, and they may have different dietary preferences which can expose them to higher risk of developing CRC.

Age is one of the significant predictors influencing the risk of developing CRC. The risk of developing

CRC increases with age. Our study is consistent with the finding, most CRC cases were in the 60-69 age group. According to Malaysia's registry, the incidence of CRC increases with age, significant at age 50 years and peak at age 75 [1, 38]. However, more recent studies have demonstrated a growing incidence of CRC among young adults age 20-49 [5]. With the rising trend of young onset CRC, the America Cancer Society has recommended to start screening the average risk population at the age of 45 years for early detection of young onset CRC among younger adults [39]. Moreover, the disease progression for CRC is long, the polyp-carcinoma progress takes approximately 10-20 years [40, 41]. Taken into account the growing trend of early onset CRC and the long period of disease progression from adenoma to carcinoma, a revision on the minimum age eligible for CRC screening on the current guidelines should be considered.

Our study revealed a significant protective effect on those who attained tertiary educational level, lowering the risk of CRC by 82% as compared to those without formal education. A study supports the finding, where lower socio-economic status and educational level has an increased risk of CRC [42]. Educational level may be indirectly related to the risk of CRC, suggestive that lower educational level may be a contributing factor towards social inequalities and health disparities, leading towards inequality in assessing healthcare services, less likely to participate in screening programs, poorer nutritional status and health knowledge [43].

There was a protective effect against CRC among those without co-morbidities. Several studies have shown that the coexistence of metabolic risk factors may have an additive effect on CRC risk. The combination of abdominal obesity, glucose intolerance and low HDL-C levels presented the highest association with the development of CRC [44, 45]. It is also clear that all NCDs including cancer are caused by the same behavioural factors such as tobacco smoking, physical inactivity, poor dietary practices and other environmental factors. Our study revealed that those who performed less vigorous physical activity had 4 times the odds of developing CRC. Therefore, between 3.5 and 4 hours of strenuous activity each week is required in order to reduce the risk of CRC [46, 47]. Preventive strategies targeting into diet and lifestyle modification will benefit the population into preventing CRC as well as other types of non-communicable diseases (NCD).

Our study demonstrated that age, gender, ethnicity, educational level, presence of comorbidities, physical activity intensity, and consumption of pickled mango and pork predicted the development of CRC. We concluded that the LR Model was a better model for CRC risk prediction with the inclusion of eight predictors as compared to the C5 Model, where LR model fared better in its overall accuracy (89.2%), discriminatory power (AUC = 0.818), sensitivity (77%) and specificity (91%). This finding is supported by several studies, in which conventional logistic regression algorithm produces a better model in cancer prediction. Logistic regression is a widely used statistical method in medical field for risk prediction [48–50]. Furthermore, as suggested by two studies, the recommended level for prediction models should demonstrate a good level of accuracy above 80% with acceptable to excellent discriminatory power. Both models in our study have demonstrated results above the recommended level [51, 52].

With this risk prediction model, risk stratification of the average risk population in Sabah can be improved by implementing a routine CRC risk stratification into daily clinical practice. Since these risk predictors are easily obtainable through patient clerking, the risk stratification can be applied during every patient consultation. That way, patients with higher risk levels could be referred for early CRC screening. Furthermore, with these known risk factors, primary prevention can be optimised. Traditionally, health awareness was given to the public by explaining that older age, presence of family history, obesity and smoking as potential risk for developing CRC, and the targeted population for health education are among those age 50 years and older [2]. Our study recommends that health education to include educating the local population about their susceptibility towards CRC, explaining which ethnicity may be more susceptible and requires an early screening, modifiable risk factors that may potentially accelerate their risk of developing CRC such as physical inactivity, poor dietary practices with frequent consumption of preserved food and red meat. Furthermore, our study recommends to start educating and risk stratifying the asymptomatic population to as young as 40 years, as the age group of 50-59 demonstrated a 3-fold increase risk of CRC based on the findings in this study, and given that the adenoma-carcinoma progress may take as long as 10-20 years to develop. Ethnic specific early CRC screening program can also be tailored to screen the high-risk ethnic groups. Our study hopes that with these risk predictors, early detection and diagnosis of CRC can be improved to downstage CRC at time of diagnosis and improve survival outcomes.

The strength of this study is that the CRC risk prediction model included risk factors that can be identified through routine data collection without involving laboratory studies, thereby enabling a more practical, less costly, and easily applicable risk stratification tool for CRC. However, the limitation is that we did not have an external test dataset to validate the prediction models. Therefore, the models should be externally validated in larger, population-based samples in future studies. Recollection bias may occur when obtaining a history of dietary practices and lifestyles using a questionnaire, particularly when participants from the case group were required to recall their habitual diet a year prior to being diagnosed with CRC. Since this is a case-control study, Furthermore, the association between BMI and CRC cannot be evaluated when BMI were obtained upon interview, as the effect of cancer and chemotherapy would have influenced the weight of the cases. Since this is a case-control study, it can be difficult to establish true temporality, as the exposures are determined after the outcome is known. Another limitation is that the study was not matched, and there were more female participants recruited as compared to male participants which did not conform to the gender proportion in Sabah. This may lead to overestimation of the effect size of association between gender and CRC. Furthermore, frequency of consumption of food items were recoded into two categories to simplify the analysis. Food quantification was not done to assess the food intake in this study. As a result, this reduces the statistical power to assess the relationship between food quantification and CRC. Dose-response between food items and cancer risk cannot be determined.

CONCLUSION

We conclude that significant risk predictors such as male gender, age group 60-69 years, Chinese ethnicity, populations with co-morbidities, physical inactivity, lower educational attainment, frequent consumption of pork and pickled mango are associated with an increased risk of developing CRC. Food preservation, such as fermentation, salting and pickling, are commonly practiced in Sabah. The frequent consumption of such foods, and their methods of preparation, should be given more attention and further observation. Different ethnic groups have demonstrated to have a disproportionate level of risk in developing CRC. Therefore, ethnicity can be a determining factor for CRC. Our study recommends the inclusion of ethnicity and dietary practices as risk predictors, as a step towards refining risk-based approaches to screen the average-risk population, and to improve early detection and diagnosis of CRC.

ACKNOWLEDGEMENT

We would like to especially thank the Director General of Health Malaysia for permission to publish this article and gratefully acknowledge the Sabah State Health Department, healthcare staff recruited as research assistants from Sikuati Health Clinic, Tandek Health Clinic, Timbua Health Clinic, Long Pasia Health Clinic, Tenom District Health Office, Inanam Health Clinic, Tawau District Health Office, Semporna District Health Office, Beluran District Health Office, Lahad Datu District Health Office, Colorectal Unit Hospital Queen Elizabeth Kota Kinabalu, and Universiti Malaysia Sabah for their support and contribution to this article. This study did not receive any specific grants from funding agencies in the public, commercial, or not-for-profit sectors. All authors declare there is no conflict of interest in preparing this article.

REFERENCES

1. Sabah State Health Department. Sabah Cancer Registry Report 2012-2016. 2019;1–27.
2. Malaysia M of H. National Strategic Plan for Colorectal Cancer (NSPCRC) 2021–2025. 2021;1–68.
3. Azizah, I.T NS, A NH, Z.A A, W M. Malaysian National Cancer Registry Report 2007-2011. 2016;1-228.
4. Khine WWT, Zhang Y, Goie GJY, Wong MS, Liong M, Lee YY, et al. Gut microbiome of pre-adolescent children of two ethnicities residing in three distant cities. Sci Rep. 2019;9(1):1–10. doi:10.1038/s41598-019-44369-y
5. Sawicki T, Ruszkowska M, Danielewicz A, Niedźwiedzka E, Arłukowicz T, Przybyłowicz KE. A Review of Colorectal Cancer in Terms of Epidemiology, Risk Factors, Development, Symptoms and Diagnosis. Cancers (Basel). 2021;13(9). doi:10.3390/cancers13092025
6. Ohri A, Robinson A, Liu B, Bhuket T, Wong R. Updated Assessment of Colorectal Cancer Incidence in the U.S. by Age, Sex, and Race/Ethnicity. Dig Dis Sci. 2020;65(6):1838–49. doi: 10.1007/s10620-019-05913-y
7. Özdemir BC, Dotto G-P. Racial Differences in Cancer Susceptibility and Survival: More Than the Color of the Skin? Trends in Cancer. 2017;3(3):181-197. doi: 10.1016/j.trecan.2017.02.002
8. Sabah State Government. Sabah's People and History. [cited 2021 Feb 23]. Available from: www.sabah.gov.my/cms/?q=en/content/people-history.
9. Renata N. Taste of Traditional Sabah Food . New Straits Times. 2017. [cited 2021 Mar 1]. Available from: www.pressreader.com/malaysia/new-straitstimes/20170413/282729111763078
10. Sapawi DKA, Ooi YBH, Ibrahim S, Lai YC. Food Safety Concerns Influence Neophobic Response Towards Kadazandusun Traditional Food in Domestic Tourists. J Tour Hosp Environ Manag. 2019;4(13):34–58.
11. WHO I. Cancers Fact Sheets : Colorectal Cancer. 2012;1–10. [cited 2021 Feb 25]. Available from: http://gco.iarc.fr/today.
12. International Agency For Research On Cancer. IARC monographs on the evaluation of carcinogenic risks to humans. IARC Monogr Eval Carcinog Risks to Humans. 2010;93:9–38.
13. Padmanabhan S, Waly MI, Taranikanti V, Guizani N, Rahman MS, Ali A, et al. Modifiable and Non-

modifiable Risk Factors for Colon and Rectal Cancer. In: Bioactive Components, Diet and Medical Treatment in Cancer Prevention. Cham: Springer International Publishing. 2018;121-130. doi: 10.1007/978-3-319-75693-6_10
14. Mohammad NMAB, Shahril MR, Shahar S, Fenech M, Sharif R. Association between Diet-related Behaviour and Risk of Colorectal Cancer: A Scoping Review. J Cancer Prev. 2022;27(4):208-220. doi: 10.15430/JCP.2022.27.4.208
15. Ministry of Health M. National Health and Morbidity Survey 2019 Technical Report, Non-Communicable Diseases: Risk Factors and other Health Problems. 2019;1-392.
16. Bujang MA, Sa'At N, Tg Abu Bakar Sidik TMI, Lim CJ. Sample Size Guidelines for Logistic Regression from Observational Studies with Large Population: Emphasis on the Accuracy Between Statistics and Parameters Based on Real Life Clinical Data. Malays J Med Sci. 2018 25(4):122. doi: 10.21315/mjms2018.25.4.12
17. IARC Working Group. Personal Habits and Indoor Combustions. IARC Monogr Eval Carcinog Risks to Humans. 2012;100E:46–167.
18. Knekt P, Ja¨rvinen R, Ja¨rvinen J, Dich J, Hakulinen T. Risk of Colorectal and Other Gastro-intestinal Cancers After Exposure To Nitrate, Nitrite and N-Nitroso Compounds: A Follow-up Study. J Cancer. 1999;80(6):852–6. doi: 10.1002/(sici)1097-0215(19990315)80:6<852::aid-ijc9>3.0.co;2-s.
19. Yan B, Zhang L, Shao Z. Consumption of processed and pickled food and esophageal cancer risk: A systematic review and meta-analysis. Bull Cancer. 2018;105(11):992–1002. doi: 10.1016/j.bulcan.2018.08.006.
20. Joshi AD, Kim A, Lewinger JP, Ulrich CM, Potter JD, Cotterchio M, Le Marchand L, Stern MC. Meat intake, cooking methods, dietary carcinogens, and colorectal cancer risk: findings from the Colorectal Cancer Family Registry. Cancer Med. 2015;4(6):936-52. doi: 10.1002/cam4.461.
21. zur Hausen H. Red meat consumption and cancer: reasons to suspect involvement of bovine infectious factors in colorectal cancer. Int J Cancer. 2012;130(11):2475-83. doi: 10.1002/ijc.27413.
22. Marques-Vidal P, Ravasco P, Camilo ME. Foodstuffs and colorectal cancer risk: A review. Clin Nutr. 2006;25(1):14–36. doi: 10.1016/j.clnu.2005.09.008.
23. Rai AK, Jeyaram K. Health benefits of functional proteins in fermented foods. In: Rai AK, Jeyaram K. Health Benefits of Fermented Foods and Beverages. 2015;(April):455–474. doi: 10.13140/RG.2.1.2335.6963
24. Awang-Kanak F, Abu Bakar MF. Traditional vegetable salad (ulam) of Borneo as source of functional food. Food Res 2020;4(1):1–12. doi:10.26656/fr.2017.4(1).138
25. Guo JY, Li X, Browning JD, Rottinghaus GE, Lubahn DB, Constantinou A, et al. Dietary Soy Isoflavones and Estrone Protect Ovariectomized ERalphaKO and Wild-Type Mice from Carcinogen-Induced Colon Cancer. J Nutr. 2004;134(1):179–82. doi: 10.1093/jn/134.1.179.
26. Yu Y, Jing X, Li H, Zhao X, Wang D. Soy isoflavone consumption and colorectal cancer risk: a systematic review and meta-analysis. Sci Rep. 2016;6:25939. doi: 10.1038/srep25939.
27. Kono S, Imanishi K, Shinchi K, Yanai F. Relationship of diet to small and large adenomas of the sigmoid colon. Jpn J Cancer Res. 1993;84(1):13-9. doi: 10.1111/j.1349-7006.1993.tb02777.x.
28. Chapelle N, Martel M, Toes-Zoutendijk E, Barkun AN, Bardou M. Recent advances in clinical practice: colorectal cancer chemoprevention in the average-risk population. Gut. 2020;69(12):2244-2255. doi: 10.1136/gutjnl-2020-320990.
29. Tasdemir SS, Sanlier N. An insight into the anticancer effects of fermented foods: A review. J Funct Foods. 2020;75(8):104281. doi: 10.1016/j.jff.2020.104281
30. Kesika P, Sivamaruthi BS, Chaiyasut C. Health promoting effects of fermented foods against cancer: an updated concise review. Food Sci Technol. 2022;42. doi: 10.1590/fst.18220
31. Gille D, Schmid A, Walther B, Vergères G. Fermented Food and Non-Communicable Chronic Diseases: A Review. Nutrients. 2018;10(4):448. doi: 10.3390/nu10040448.
32. Young GL, Ki WL, Ji YK, Kyoung HK, Hyong JL. Induction of apoptosis in a human lymphoma cell line by hydrophobic peptide fraction separated from anchovy sauce. BioFactors. 2004;21(1–4):63–7. doi: 10.1002/biof.552210112.
33. Sinha R, Peters U, Cross AJ, Kulldorff M, Weissfeld JL, Pinsky PF, Rothman N, Hayes RB. Meat, meat cooking methods and preservation, and risk for colorectal adenoma. Cancer Res. 2005;65(17):8034-41. doi: 10.1158/0008-5472.CAN-04-3429.
34. Tuan N, Giang L, Ngoan L. Case-Control Study of Diet, Cooking Methods and Colorectal Cancer In Viet Nam. Asian Pacific J. Cancer Prev. 2018;21(8):2217-2223. doi: 10.31557/APJCP.2020.21.8.2217
35. World Health Organisation. Cancer: Carcinogenicity of the consumption of red meat and processed meat. [cited 2021 Jul 23]. Available from: www.who.int/news-room/q-a-detail/cancer-carcinogenicity-of-the-consumption-of-red-meat-and-processed-meat.
36. Khor S, Haupt EC, Hahn EE, Lindsay ;, Lyons JL, Shankaran V, et al. Racial and Ethnic Bias in Risk Prediction Models for Colorectal Cancer Recurrence When Race and Ethnicity Are Omitted as Predictors. JAMA Netw Open. 2023;6(6):e2318495. doi: 10.1001/jamanetworkopen.2023.18495.

37. Theuer C. A Call To Reinstate Race And Ethnicity Based Colorectal Cancer Screening Guidelines. [cited 2023 June 21]. Available from: www.forbes.com/sites/forbesmarketplace/2023/03/14/a-call-to-reinstate-race-and-ethnicity-based-colorectal-cancer-screening-guidelines/?sh=53bb13646ddd
38. National Cancer Institute. Malaysia National Cancer Registry Report (MNCRR) 2012-2016. 2019;1-116.
39. Wolf AMD, Fontham ETH, Church TR, Flowers CR, Guerra CE, LaMonte SJ, Etzioni R, McKenna MT, Oeffinger KC, Shih YT, Walter LC, Andrews KS, Brawley OW, Brooks D, Fedewa SA, Manassaram-Baptiste D, Siegel RL, Wender RC, Smith RA. Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. CA Cancer J Clin. 2018;68(4):250-281. doi: 10.3322/caac.21457.
40. Rawla P, Sunkara T, Barsouk A. Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. Prz Gastroenterol. 2019;14(2):89-103. doi: 10.5114/pg.2018.81072.
41. Stryker SJ, Wolff BG, Culp CE, Libbe SD, Ilstrup DM, MacCarty RL. Natural history of untreated colonic polyps. Gastroenterology. 1987;93(5):1009-13. doi: 10.1016/0016-5085(87)90563-4.
42. Aarts MJ, Lemmens VE, Louwman MW, Kunst AE, Coebergh JW. Socioeconomic status and changing inequalities in colorectal cancer? A review of the associations with risk, treatment and outcome. Eur J Cancer. 2010;46(15):2681-95. doi: 10.1016/j.ejca.2010.04.026.
43. Ishaq SL, Rapp M, Byerly R, McClellan LS, O'Boyle MR, Nykanen A, Fuller PJ, Aas C, Stone JM, Killpatrick S, Uptegrove MM, Vischer A, Wolf H, Smallman F, Eymann H, Narode S, Stapleton E, Cioffi CC, Tavalire HF. Framing the discussion of microorganisms as a facet of social equity in human health. PLoS Biol. 2019;17(11):e3000536. doi: 10.1371/journal.pbio.3000536.
44. Cho YK, Lee J, Kim HS, Park JY, Lee WJ, Kim YJ, Jung CH. Metabolic health is a determining factor for incident colorectal cancer in the obese population: A nationwide population-based cohort study. Cancer Med. 2021;10(1):220-229. doi: 10.1002/cam4.3607.
45. Zhang C, Cheng Y, Luo D, Wang J, Liu J, Luo Y, Zhou W, Zhuo Z, Guo K, Zeng R, Yang J, Sha W, Chen H. Association between cardiovascular risk factors and colorectal cancer: A systematic review and meta-analysis of prospective cohort studies. EClinicalMedicine. 2021;34:100794. doi: 10.1016/j.eclinm.2021.100794.
46. Cerin E, Leslie E, Bauman A, Owen N. Levels of physical activity for colon cancer prevention compared with generic public health recommendations: population prevalence and sociodemographic correlates. Cancer Epidemiol Biomarkers Prev. 2005(4):1000-2. doi: 10.1158/1055-9965.EPI-04-0622.
47. Slattery ML. Physical activity and colorectal cancer. Sports Med. 2004;34(4):239-52.
doi:10.2165/00007256-200434040-00004. PMID: 15049716.
48. Feng CH, Disis ML, Cheng C, Zhang L. Multimetric feature selection for analyzing multicategory outcomes of colorectal cancer: random forest and multinomial logistic regression models. Lab Investig. 2022;102(3):236–44. doi: 10.1038/s41374-021-00662-x.
49. Li W, Zhao LZ, Ma DW, Wang DZ, Shi L, Wang HL, et al. Predicting the risk for colorectal cancer with personal characteristics and fecal immunochemical test. Med (United States). 2018;97(18):1–7. doi: 10.1097/MD.0000000000010529.
50. Rabinovitch-Chable H, Durand K, Genet D, Marin B, Dzugan H, Léobon S, et al. Logistic regression model of the clinical response to 5-fluorouracil based chemotherapy for metastatic colorectal cancer patients. Anticancer Res. 2006;26(5B):3885–92. PMID: 17094418.
51. Hosmer DW, Lemeshow S. Interpretation of the Fitted Logistic Regression Model. In: Hosmer DW, Lemeshow S. Applied Logistic Regression. New York: John Wiley and Sons, Inc. 2000:47-90.
52. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology. 1982;143(1):29–36. doi: 10.1148/radiology.143.1.7063747.