

Risk prediction models for colorectal cancer: a systematic review

Juliet A Usher-Smith¹, Fiona M Walter^{1,2}, Jon D Emery^{1,2}, Aung K Win³, Simon J Griffin¹

¹ The Primary Care Unit, Institute of Public Health, University of Cambridge School of Clinical Medicine, Strangeways Research Laboratory, 2 Wort's Causeway, Cambridge, CB1 8RN, UK

² Department of General Practice, Melbourne Medical School Faculty of Medicine, Dentistry & Health Sciences The University of Melbourne, 200 Berkeley Street, Carlton, Victoria, 3053, Australia

³ Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, Level 4, 207 Bouverie Street, The University of Melbourne, Victoria, 3010, Australia

Running title: Risk prediction models for colorectal cancer

Key words: Risk, prediction, colorectal cancer, review

Financial support: J Usher-Smith is funded by a National Institute of Health Research (NIHR) Clinical Lectureship and F Walter by an NIHR Clinician Scientist award. J Emery is funded by an Australian National Health and Medical Research Council (NHMRC) Practitioner Fellowship. A Wong has an NHMRC Early Career Fellowship. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. All researchers were independent of the funding body and the study sponsors and funder had no role in study design; data collection, analysis and interpretation of data; in the writing of the report; or decision to submit the article for publication.

Correspondence to: J Usher-Smith, The Primary Care Unit, University of Cambridge, Strangeways Research Laboratory, 2 Worts Causeway, Cambridge CB1 8RN. Email: jau20@medschl.cam.ac.uk Tel: 01223 748693 Fax: 01223 768412

Conflicts of interest: None

Word count: 3527

Total number of figures and tables: 5

ABSTRACT

Colorectal cancer (CRC) is the second leading cause of cancer-related death in Europe and the United States. Survival is strongly related to stage at diagnosis and population-based screening reduces CRC incidence and mortality. Stratifying the population by risk offers the potential of improving the efficiency of screening. In this systematic review we searched Medline, EMBASE and the Cochrane Library for primary research studies reporting or validating models to predict future risk of primary CRC for asymptomatic individuals. 12,808 papers were identified from the literature search and nine through citation searching. 52 risk models were included. Where reported ($n=37$), half the models had acceptable-to-good discrimination ($c\text{-statistic}>0.7$) in the derivation sample. Calibration was less commonly assessed ($n=21$), but overall acceptable. In external validation studies, 10 models showed acceptable discrimination ($c\text{-statistic } 0.71\text{-}0.78$). These include two with only three variables (age, gender and BMI; age, gender and family history of CRC). A small number of prediction models developed from case-control studies of genetic biomarkers also show some promise but require further external validation using population-based samples. Further research should focus on the feasibility and impact of incorporating such models into stratified screening programmes.

INTRODUCTION

Colorectal cancer (CRC) is the second leading cause of cancer-related death in Europe and the United States(1). Survival is strongly related to stage at diagnosis(2) and population-based screening has been shown to significantly reduce CRC incidence and mortality(3–6). Stratifying the population into risk categories offers the potential to improve the efficiency of this screening by tailoring the intensity of screening, or preventive approaches, to the predicted level of risk. Providing patients and practitioners with a personalised risk assessment may also encourage engagement in risk reducing behaviours, including participation in screening or prevention programmes and lifestyle changes to reduce incidence of disease(7).

A number of risk prediction models for CRC have been developed and two previous reviews of these have been published(8, 9). However, neither was comprehensive, and since those reviews were published several new risk models have been developed. This paper provides the first comprehensive analysis of risk prediction tools for risk of primary colorectal cancer in asymptomatic individuals within the general population. It includes analysis of the range of 87 variables in addition to genes and SNPs included in each model, the predictive ability of the different risk models and their potential applicability and practical use for population based stratification.

MATERIALS AND METHODS

We performed a systematic literature review following an a priori established study protocol (available on request).

Search strategy

We performed an electronic literature search of Medline, EMBASE and the Cochrane Library from Jan 2000 up to March 2014 with no language limits using a combination of subject headings incorporating ‘colorectal cancer’, ‘risk/risk factor/risk assessment/chance’ and ‘prediction/model/score’ (see Supplementary File 1 for complete search strategy for Medline and EMBASE). We then manually screened the reference lists of all included papers.

Study selection

We included studies if they fulfilled all of the following criteria: (i) published as a primary research paper in a peer-reviewed journal; (ii) identify risk factors for developing colon, rectal or colorectal cancer or advanced colorectal neoplasia at the level of the individual; (iii) provide a measure of relative or absolute risk using a combination of two or more risk factors that allows identification of people at higher risk of colon and/or rectal cancer; and (iv) are applicable to the general population. Studies including only highly selected groups, for example immunosuppressed patients, organ transplant recipients or those with a previous history of colon and/or rectal cancer were excluded. Conference proceedings were also excluded after contacting the authors to confirm the results had not been published elsewhere in a peer-reviewed journal.

One reviewer (JUS) performed the search and screened the titles and abstracts to exclude papers that were clearly not relevant. Two reviewers (FW and SG) independently assessed a random selection of 5% of the papers each. The full text was examined where a definite decision to reject could not be made based on title and abstract alone. At least two reviewers (JUS and FW/SG/JE) independently assessed all full-text papers, and those deemed not to meet inclusion criteria by both researchers were excluded. We discussed papers for which it was unclear whether or not the inclusion criteria were met at consensus meetings.

Papers written in languages other than English were translated into English for assessment and subsequent data extraction.

Data extraction and synthesis

Data were extracted independently by at least two researchers (JUS and FW/SG/JE) using a standardised form to minimise bias. The form included details on: (i) the development of the model, including potential risks of bias such as the study design, selection of participants, and the variables considered for inclusion in the model and how they were selected; (ii) the risk model itself, including the variables included and requirement for data collection; (iii) the performance of the risk model in the development population; and (iv) any validation studies of the risk model and/or data collection tool, including the study design and performance of the risk model. In this process the methods of studies published for each risk model were classified according to the TRIPOD guidelines(10) and tabulation of the methods allowed assessment of bias. For studies which included multiple different models, for example separate models for men and women or for self-assessment and physician assessment, all were included separately.

RESULTS

Identified risk models

After duplicates were removed, the search identified 12,808 papers. Of these, 12,727 were excluded at title and abstract level and a further 50 after full-text assessment. After title and abstract screening by the first reviewer (JUS), no additional papers met the inclusion criteria in the random 10% screened by a second reviewer (FW/SG). The most common reasons for exclusion at full-text level were that the papers included symptomatic populations, were conference abstracts or did not include a risk score (Figure 1). Four were

excluded as they included circulating biomarkers that were felt to detect prevalent undiagnosed disease rather than estimate future risk(11–14).

Nine further papers were identified through citation searching, giving 40 papers describing 52 risk models for inclusion in the analysis and 6 external validation studies(15–20). Table 1 summarises these 52 risk models. Thirteen have advanced colonic neoplasia (defined as invasive cancer, an adenoma 10mm or more, a villous adenoma (at least 25% villous), or an adenoma with high grade dysplasia) as the outcome(21–32), 13 colon cancer(33–41), 20 colorectal cancer (CRC)(31, 36, 38, 39, 41–54), and 6 rectal cancer(37–39). Most include both men and women, but 16 are specific to either men or women. Six include only variables that are available in routine medical records. The majority ($n=32$) include variables obtained via a self-completed questionnaire. These range from questionnaires with only one or two simple questions concerning family history(26, 27, 32, 50, 52), diet(44) or physical activity(38) to those including detailed dietary habits, aspirin/NSAID use, oestrogen and HRT use, inflammatory bowel disease, previous colonoscopy or sigmoidoscopy and polyp history and the most complex including 15 variables(35). Six, all from the same study, use data from a self-completed questionnaire and results of blood tests for fasting plasma glucose and total cholesterol(39), four a blood test alone for genetic biomarkers(45, 48, 49, 51), and four a self-completed questionnaire and genetic biomarkers(43, 54, 46). Between them, the authors of the 52 risk models considered 87 different risk factors (Table 2).

Development of the risk models

Further details of the development of each model and the risks of bias are given in Supplementary Tables 1a-d. Seventeen were developed from case-control studies with cases identified from hospitals or cancer disease registries and controls from primary care ($n=1$),

hospitals ($n=5$), other research studies ($n=2$), random-digit dialling ($n=7$), spouses ($n=1$), healthy individuals or blood donors ($n=1$). Seventeen were developed from cohort studies with between 21,581 and 1,326,058 participants and most identifying cases of cancer through cancer registries over a 10-20 year follow-up period. Fourteen were cross-sectional studies of participants attending for screening colonoscopy and all but one had advanced colorectal neoplasia as the outcome. Three risk models were developed from a review of the literature(35), a meta-analysis of risk factors(47) or modelling in a simulated population(54).

Discrimination of the risk models

The performance of 42 of the 52 models was reported in at least one of either the development population ($n=31$), using bootstrapping or cross-validation ($n=13$), a subset of the initial development population ($n=3$), or an external population ($n=21$). Details of the discrimination, calibration and accuracy are given in Table 3 and details of the methods for those using a subset of the initial population or external populations in Supplementary Table 2.

Discrimination, as measured by the area under the receiver operator curve (AUROC), was reported for 37 of the risk models, and these values are summarised in Figure 2 in which the models are grouped into five groups according to the type of variables included (routine data only, self-completed questionnaire, self-completed questionnaire and non-genetic biomarkers, genetic biomarkers, and self-completed questionnaires plus genetic biomarkers). Within each group the models are order according to the number of variables included. The models on the left are, therefore, those with the fewest and most easily obtained variables and the more complex models are towards the right of the figure. Most models have acceptable to good discrimination with AUROCs between 0.65 and 0.75.

Amongst those models including only routinely available data, the best performing and validated model for advanced colorectal neoplasia was developed by Betes *et al* among 2,210 asymptomatic individuals attending routine CRC screening in Spain(21). It includes only age, gender and BMI and has AUROCs of 0.65 (95%CI: 0.61-0.69) and 0.71 (95%CI: 0.64-0.78) in external validation studies in China(22, 24). The only risk scores using routine data for colon cancer and CRC were developed by Driver *et al* from a cohort of 21,581 men in the USA(36). The score for colon cancer includes age, BMI and history of smoking and has an AUROC of 0.72 in that population and the score for CRC includes those variables plus alcohol consumption and has similar discrimination in bootstrap analysis (AUROC=0.69)(36).

The second group of risk models used self-completed questionnaire and, as illustrated by the absence of any clear trend in the AUROC within that group in Figure 2, there is no clear improvement in discrimination as increasing numbers of variables are added from self-completed questionnaires to routine data. There is a suggestion from the third group of risk models that adding fasting serum glucose and total cholesterol to self-completed questionnaire variables might improve the discrimination in the scores developed by Shin *et al* using a South Korean population of men(39), but this same improvement above other risk models containing only routine or questionnaire data was not seen in women.

The two models with the highest discrimination are both in the group based entirely on genetic biomarkers and were developed from small case-control studies. The model by Han *et al* 2008 includes 5 genes (BANK1, B-cell scaffold protein with ankyrin repeats 1; BCNP1, B-cell novel protein 1; CDA, cytidine deaminase; MGC20553, FERM domain containing 3; MS4A, membrane-spanning 4 domains) identified from a case-control study including 58 patients with CRC and 57 disease-free controls using hierarchical cluster analysis and logistic regression(45). In that development population the biomarker panel has

an AUROC of 0.88 (95%CI: 0.81-0.94). It has yet to be externally validated. The model developed by Marshall *et al* 2010 includes seven genes (ANXA3, Annexin A3; CLEC4D, C-type lectin domain family4, member D; IL2RB, Interleukin 2 receptor, beta; LMNB1, Lamin B1; PRRG4, Proline risk Gla 4; TNFAIP6, tumour necrosis factor, alpha-induced protein 6; VNN1, Vanin 1) similarly identified from a case-control study with 112 patients with CRC and 120 disease-free controls from hospitals in Canada and the USA(49). In that population the model has an AUROC of 0.80 (95%CI: 0.74-0.85) and in a separate sample of 99 patients with CRC and 111 controls in Malaysia the AUROC was reported as 0.76 (95%CI: 0.70-0.82). The third risk model based entirely on genetic biomarkers also has acceptable discrimination. It was developed by Wang *et al* in Taiwan, again from a case-control study, and includes 16 SNPs from a GWAS study in Asian people(51). It has an AUROC of 0.77 in the development population and 0.72 in cross-validation.

The final group of four risk models including both genetic biomarkers and phenotypic variables, however, do not have such good discrimination and adding variable numbers of different SNPs to data available from self-completed questionnaires does not appear to improve discrimination. The addition of 10 SNPs to age, gender and family history(43) or three SNPs to age and family history(46) in case-control studies does not improve discrimination over age, gender and family history alone(50) (AUROC 0.57 and 0.73 (95%CI: 0.68-0.77) (male), 0.65 (95%CI: 0.62-0.68) (female) compared to 0.67). The discrimination of a model with 14 SNPs added to BMI, smoking, alcohol, fibre intake, red meat intake and physical activity(54) has an AUROC of 0.63 in a simulated population which is no better than those models using only routinely available data in cross-sectional or cohort studies.

Calibration of the risk models

Calibration was reported for 21 of the 52 models. In most cases it was reported as the Hosmer-Lemeshow statistic ($n=9$) or Chi-squared test ($n=6$) with p values ranging from $p=0.0003$ to $p=0.94$. Where expected:observed ratios were given the confidence intervals all cross one except for the model by Ma *et al* 2010 for colon cancer where it is 1.19 (95%CI: 1.03-1.37)(38).

Sensitivity and specificity of the risk models

Sensitivity and specificity were reported for only seven models. Two of these were the genetic models developed by Han and Marshall which have sensitivities of 88% and 71.7% and specificities of 64% and 71.2% in external populations respectively(45, 49). The other five were all risk models for advanced colorectal neoplasia and range from high sensitivity (92.4%) and low specificity (13.9%) in Kaminski(26) to low sensitivity (40%) and high specificity (93%) in Stegeman(29).

Comparison of different outcomes

Five studies(31, 36, 38, 37, 39) simultaneously developed risk models for more than one of advanced colorectal neoplasia, CRC, colon cancer and rectal cancer. All showed that beta-coefficients and included variables differed slightly between different sites but only two provided any comparative data. Tao reported the performance of the same model for predicting advanced colorectal neoplasia or CRC and showed that the discrimination was similar (AUROC 0.68 for advanced colorectal neoplasia and 0.66 for CRC)(31). Driver showed that the AUROC of a predictive model developed for colon cancer was only slightly superior to the model developed for CRC when predicting CRC risk (0.717 vs 0.695), but the goodness-of-fit test showed it to perform less well than the CRC model (Hosmer-Lemeshow statistic 0.43 vs 0.91)(36).

DISCUSSION

Principal findings

To our knowledge this is the first comprehensive systematic review of risk prediction models for CRC. It shows that multiple risk models exist for predicting the risk of developing CRC, colon cancer, rectal cancer or advanced colorectal neoplasia in asymptomatic populations, and that they have the potential to identify individuals at high risk of disease. The discrimination of the models, as measured by AUROC, compare favourably with risk models used for other cancers, including breast cancer(55) and melanoma(56), and several include only variables recorded in routine medical records and so could be implemented into practice without the need for further data collection. Grouping risk models according to type and number of variables included (Figure 2) also shows that there is no clear improvement in discrimination as increasing numbers of variables are added from self-completed questionnaires to routine data, or in studies in which genetic biomarkers are added to data from self-completed questionnaires. A small number of risk models developed from case-control studies of genetic biomarkers alone show some promise but require further external validation in population-based samples.

Strengths and weaknesses

The main strength of this review is our use of a broad search strategy and careful screening of possible papers for inclusion. Whilst we cannot exclude publication bias or the possibility that there are other risk models that we did not identify, using this systematic approach enabled us to identify over 3 times as many risk models as reported in previous reviews in this area(8, 9). This review is, therefore, the most comprehensive to date and the inclusion of less well cited risk models allows us to demonstrate for the first time the relative

performance of simple and more complex models. However, as we included only those risk models applicable to asymptomatic individuals from the general population, these models are not applicable to those with familial syndromes, such as Lynch syndrome or familial adenomatous polyposis, or those with existing cancer. Most of the risk models were developed from predominantly white populations in Europe or America or Asian populations in China, Japan, Taiwan and Korea, with only two from Arabic countries and none from Australasia. There is a well-recognized high degree of heterogeneity by nationality in CRC incidence with an up to 10-fold difference internationally(57). Much of this variation is thought to be due to differences in environmental risk factors as the incidence rate of CRC in migrants approaches that of the host country within one or two generations(58). The risk models in this review may, therefore, be less applicable to these less well represented populations.

Implications for clinicians and policy makers

There is now substantial evidence that the incidence of, and mortality from, CRC can be reduced by screening with faecal occult blood testing(59–61), flexible sigmoidoscopy(62, 63), or colonoscopy(64–66), and multiple economic analyses support the cost-effectiveness of population-based CRC screening(67–69). This review shows that risk models exist that have the potential to stratify the general population into risk categories and allow screening and preventive strategies to be targeted at those most likely to benefit whilst leaving those at low risk of disease unexposed to direct and indirect harms of screening programmes. This might improve the cost-effectiveness of CRC screening(70) and would address concerns about demand and capacity for colonoscopy(71, 72). It would also provide an opportunity to implement potential chemo-preventive medications such as non-steroidal anti-inflammatory drugs. These drugs are currently not recommended for asymptomatic adults at average risk

for CRC(73), but both the United States Preventive Services Task Force(74) and a recent international consensus panel(75) advocate additional research into the use of aspirin in high-risk individuals for whom benefits might outweigh the harms. The use of risk prediction models would also potentially increase uptake of screening and provide an opportunity to give information to encourage lifestyle changes. Despite the known mortality benefit of CRC screening, large numbers of eligible people do not participate in CRC screening programs(76, 77). Whilst the reasons for non-participation are complex, several studies have suggested that high-risk individuals are more likely to be up-to-date with CRC screening and adhere to physician recommendations(77–80). Knowledge of their risk, both within or outside screening programmes, may also encourage adoption of more healthy lifestyles which might further improve outcomes: it is estimated that between 30% and 70% of the overall burden of colon cancers in the US and UK populations could be prevented through moderate changes in diet and lifestyle(81, 82), and information about individualised colon cancer risk has been shown to lead to a reduction in multiple behavioural risk factors in patients with a history of colon adenoma(83).

Several barriers, however, exist to the incorporation of risk prediction models into practice. The main one is the practical challenge of collecting the necessary risk factor information. Many of the included risk scores used data collected from food frequency questionnaires. Whilst this allows accurate estimates for research, their application is unlikely to be practical at the population level. Similarly, risk scores including genetic biomarkers require sample collection and processing and some means of feeding back results to individuals. Although from figure 2 it appears as if the two models with the highest reported discrimination were both based on genes, these were developed in small case-control studies which will tend to over-estimate performance in the general population. Several risk models including genetic biomarkers also performed no better than those based on routine

information and GWA studies of colorectal cancer have shown that the CRC risks associated with each of the variants are at best modest (relative risks of 1.1-1.3), with the distribution of risk alleles following a normal distribution in both CRC cases and controls(84). As our understanding of these genetic biomarkers increases, and point-of-care genetic profiling becomes more widely available, more accurate models incorporating genomic markers will become easier to implement. A risk model that is able to predict CRC, colon cancer, rectal cancer and advanced colorectal neoplasia would also clearly have more utility in the clinical setting than separate models for each and this review also shows that to be possible: where studies developed separate risk models for different sites, the final models did include different variables, but these differences tended to be small and the performance of the models similar(31, 36).

Unanswered questions and future research

Whilst the potential clinical and economic benefits of successfully integrating a risk prediction model for CRC into clinical practice could be substantial, it remains to be defined what role the currently available and emerging models can have in practice and a number of steps are required to establish a viable useable risk profile. Firstly, this review provides comparative data on the performance of existing risk models but ideally the choice of risk model for each country would be based on validation studies in each population of interest(10). Further studies are therefore needed to compare the performance of these risk models, including those for different sites, simultaneously in large cohorts. This is particularly the case for those risk models incorporating genetic biomarkers which have mostly been developed using small case-control studies and which may perform substantially less well in population-based studies. Secondly, research is needed to establish the optimal implementation strategies. This includes modelling studies comparing the impact on

morbidity and mortality and cost-effectiveness of different implementation strategies in comparison to current programmes based on age alone and consensus meetings with expert groups. Thirdly, qualitative research with members of the public and practitioners is needed to determine how best to communicate the risk output and to assess the feasibility, acceptability of any risk based programme. Finally, before any risk model is introduced into routine clinical practice, implementation studies, ideally randomised controlled trials, are needed to assess the benefits and potential adverse consequences of applying these models in practice.

ACKNOWLEDGEMENTS

We thank Isla Kuhn, Reader Services Librarian, University of Cambridge Medical Library, for her help developing the search strategy.

REFERENCES

1. Ferlay, J, Soerjomataram I, I, Dikshit, R, Eser, S, Mathers, C, Rebelo, M, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* 2014; 136: E359–86.

Cancer Research UK (2009). *By stage at diagnosis. London: Cancer Research UK.*

3. Hardcastle, JD, Chamberlain, JO, Robinson, MH, Moss, SM, Amar, SS, Balfour, TW, et al. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. *Lancet* 1996; 348: 1472–7.

4. Holme, Ø, Bretthauer, M, Fretheim, A, Odgaard-Jensen, J, Hoff, G. Flexible sigmoidoscopy versus faecal occult blood testing for colorectal cancer screening in asymptomatic individuals. *Cochrane database Syst. Rev.* 2013; 9: CD009259.
5. Kronborg, O, Fenger, C, Olsen, J, Jørgensen, OD, Søndergaard, O. Randomised study of screening for colorectal cancer with faecal-occult-blood test. *Lancet* 1996; 348: 1467–71.
6. Lindholm, E, Brevinge, H, Haglund, E. Survival benefit in a randomized clinical trial of faecal occult blood screening for colorectal cancer. *Br. J. Surg.* 2008; 95: 1029–36.
7. Hall, WD, Mathews, R, Morley, KI. Being more realistic about the public health impact of genomic medicine. *PLoS Med.* 2010; 7: e1000347.
8. Win, AK, Macinnis, RJ, Hopper, JL, Jenkins, M a. Risk prediction models for colorectal cancer: a review. *Cancer Epidemiol. Biomarkers Prev.* 2012; 21: 398–410.
9. Ma, GK, Ladabaum, U. Personalizing Colorectal Cancer Screening: A Systematic Review of Models to Predict Risk of Colorectal Neoplasia. *Clin. Gastroenterol. Hepatol.* 2014; 12: 1624–34.e1.
10. Collins, GS, Reitsma, JB, Altman, DG, Moons, KGM. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): The TRIPOD Statement. *Ann. Intern. Med.* 2015; 162: 55–63.
11. Villadiego-Sánchez, JM, Ortega-Calvo, M, Pino-Mejías, R, Cayuela, a., Iglesias-Bonilla, P, García-de La Corte, F, et al. Multivariate explanatory model for sporadic carcinoma of the colon in Dukes' stages I and IIa. *Int. J. Med. Sci.* 2009; 6: 43–50.

12. García-Bilbao, A, Armañanzas, R, Ispizua, Z, Calvo, B, Alonso-Varona, A, Inza, I, et al. Identification of a biomarker panel for colorectal cancer diagnosis. *BMC Cancer* 2012; 12: 43.
13. Nishiumi, S, Kobayashi, T, Ikeda, A, Yoshie, T, Kibi, M, Izumi, Y, et al. A novel serum metabolomics-based diagnostic approach for colorectal cancer. *PLoS One* 2012; 7: 1–10.
14. Pengjun, Z, Xinyu, W, Feng, G, Xinxin, D, Yulan, L, Juan, L, et al. Multiplexed cytokine profiling of serum for detection of colorectal cancer. *Future Oncol.* 2013; 9: 1017–27.
15. Emmons, K., Koch-Weser, S, Atwood, K, Conboy, L, Rudd, R, Colditz, G. A Qualitative Evaluation of the Harvard Cancer Risk Index. *J. Health Commun.* 1999; 4: 181–193.
16. Emmons, KM, Wong, M, Puleo, E, Weinstein, N, Fletcher, R, Colditz, G. Tailored computer-based cancer risk communication: correcting colorectal cancer risk perception. *J. Health Commun.* 9: 127–41.
17. Kim, DJ, Rockhill, B, Colditz, G a. Validation of the Harvard Cancer Risk Index: A prediction tool for individual cancer risk. *J. Clin. Epidemiol.* 2004; 57: 332–340.
18. 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.
19. Schroy, PC, Coe, AM, Mylvaganam, SR, Ahn, LB, Lydotes, MA, Robinson, PA, et al. The Your Disease Risk Index for Colorectal Cancer Is an Inaccurate Risk Stratification Tool for Advanced Colorectal Neoplasia at Screening Colonoscopy. *Cancer Prev. Res.* 2012; 5: 1044–1052.

20. Yip, K-T, Das, PK, Suria, D, Lim, C-R, Ng, G-H, Liew, C-C. A case-controlled validation study of a blood-based seven-gene biomarker panel for colorectal cancer in Malaysia. *J. Exp. Clin. Cancer Res.* 2010; 29: 128.
21. Betés, M, Muñoz-Navas, M a., Duque, JM, Angós, R, Macías, E, Súbtil, JC, et al. Use of Colonoscopy as a Primary Screening Test for Colorectal Cancer in Average Risk People. *Am. J. Gastroenterol.* 2003; 98: 2648–2654.
22. Cai, QC, Yu, E Da, Xiao, Y, Bai, WY, Chen, X, He, LP, et al. Derivation and validation of a prediction rule for estimating advanced colorectal neoplasm risk in average-risk chinese. *Am. J. Epidemiol.* 2012; 175: 584–593.
23. Chen, G, Mao, B, Liu, Q, Qian, J, Liu, L. Derivation and validation of a prediction rule for estimating colorectal neoplasm risk in asymptomatic individuals in southern Jiangsu province. *World J. Gastroenterol.* 2013; 21: 4043–4049.
24. Chen, G, Mao, B, Pan, Q, Liu, Q, Xu, X, Ning, Y. Prediction rule for estimating advanced colorectal neoplasm risk in average-risk populations in southern Jiangsu Province. 2014; 26: 4–11.
25. Hassan, C, Pooler, BD, Kim, DH, Rinaldi, A, Repici, A, Pickhardt, PJ. Computed tomographic colonography for colorectal cancer screening: Risk factors for the detection of advanced neoplasia. *Cancer* 2013; 119: 2549–2554.
26. Kaminski, MF, Polkowski, M, Kraszewska, E, Rupinski, M, Butruk, E, Regula, J. A score to estimate the likelihood of detecting advanced colorectal neoplasia at colonoscopy. *Gut* 2014; 63: 1112–9.

27. Lin, OS, Kozarek, R a., Schembre, DB, Ayub, K, Gluck, M, Cantone, N, et al. Risk Stratification for Colon Neoplasia: Screening Strategies Using Colonoscopy and Computerized Tomographic Colonography. *Gastroenterology* 2006; 131: 1011–1019.
28. Lin, Y, Yu, M, Wang, S, Chappell, R, Imperiale, TF. Advanced colorectal neoplasia risk stratification by penalized logistic regression. *Stat. Methods Med. Res.* 2013; 00000: 0–15.
29. Stegeman, I, Wijkerslooth, TR de, Stoop, EM, Leerdam, ME van, Dekker, E, Ballegooijen, M van, et al. Combining risk factors with faecal immunochemical test outcome for selecting CRC screenees for colonoscopy. *Gut* 2014; 63: 466–71.
30. Stegeman, I, Wijkerslooth, TR de, Stoop, EM, Leerdam, ME van, Dekker, E, Ballegooijen, M van, et al. Colorectal cancer risk factors in the detection of advanced adenoma and colorectal cancer. *Cancer Epidemiol.* 2013; 37: 278–83.
31. Tao, S, Hoffmeister, M, Brenner, H. Development and Validation of a Scoring System to Identify Individuals at High Risk for Advanced Colorectal neoplasms Who Should Undergo Colonoscopy Screening. *Clin. Gastroenterol. Hepatol.* 2014; 12: 478–485.
32. Yeoh, K-G, Ho, K-Y, Chiu, H-M, Zhu, F, Ching, JYL, Wu, D-C, et al. The Asia-Pacific Colorectal Screening score: a validated tool that stratifies risk for colorectal advanced neoplasia in asymptomatic Asian subjects. *Gut* 2011; 60: 1236–1241.
33. Almurshed, KS. Colorectal cancer: Case-control study of sociodemographic, lifestyle and anthropometric parameters in Riyadh. *East. Mediterr. Heal. J.* 2009; 15: 817–826.

34. Camp, NJ, Slattery, ML. Classification tree analysis: a statistical tool to investigate risk factor interactions with an example for colon cancer (United States). *Cancer Causes Control* 2002; 13: 813–823.
35. Colditz, G a, Atwood, K a, 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 Control* 2000; 11: 477–488.
36. Driver, J a, Gaziano, JM, Gelber, RP, Lee, I-M, Buring, JE, Kurth, T. Development of a risk score for colorectal cancer in men. *Am. J. Med.* 2007; 120: 257–63.
37. Freedman, a. N, 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.* 2009; 27: 686.
38. Ma, E, Sasazuki, S, Iwasaki, M, Sawada, N, Inoue, M. 10-Year risk of colorectal cancer: development and validation of a prediction model in middle-aged Japanese men. *Cancer Epidemiol.* 2010; 34: 534–41.
39. Shin, A, Joo, J, Yang, H-R, Bak, J, Park, Y, Kim, J, et al. Risk prediction model for colorectal cancer: national health insurance corporation study, Korea. *PLoS One* 2014; 9: e88079.
40. Wei, EK, Colditz, G a., Giovannucci, EL, Fuchs, CS, Rosner, B a. Cumulative risk of colon cancer up to age 70 years by risk factor status using data from the nurses' health study. *Am. J. Epidemiol.* 2009; 170: 863–872.

41. Wei, EK, Giovannucci, E, Wu, K, Rosner, B, Charles, S, Willett, WC, et al. Comparison of Risk Factors for Colon and rectal Cancer. *Int. J. Cancer* 2004; 108: 433–442.
42. Bener, A, Moore, M a., Ali, R, Ayoubi, HR El. Impacts of family history and lifestyle habits on colorectal cancer risk: A case-control study in Qatar. *Asian Pacific J. Cancer Prev.* 2010; 11: 963–968.
43. Dunlop, MG, Tenesa, a., Farrington, SM, Ballereau, S, Brewster, DH, Koessler, T, et al. Cumulative impact of common genetic variants and other risk factors on colorectal cancer risk in 42 103 individuals. *Gut* 2012; 871–881.
44. Guesmi, F, Zoghalmi, A, Sghaiier, D, Nouira, R, Dziri, C. Alimentary factors promoting colorectal cancer risk: A prospective epidemiological study. *Tunis. Med.* 2010; 88: 184–189.
45. Han, M, Choong, TL, Hong, WZ, Chao, S, Zheng, R, Kok, TY, et al. Novel blood-based, five-gene biomarker set for the detection of colorectal cancer. *Clin. Cancer Res.* 2008; 14: 455–460.
46. Jo, J, Nam, CM, Sull, JW, Yun, JE, Kim, SY, Lee, SJ, et al. Prediction of Colorectal Cancer Risk Using a Genetic Risk Score: The Korean Cancer Prevention Study-II (KCPS-II). *Genomics Inform.* 2012; 10: 175.
47. Johnson, CM, Wei, C, Ensor, JE, Smolenski, DJ, Amos, CI, Levin, B, et al. Meta-analyses of colorectal cancer risk factors. *Cancer Causes Control* 2013; 24: 1207–22.
48. Lubbe, SJ, Bernardo, MC Di, Broderick, P, Chandler, I, Houlston, RS. Comprehensive evaluation of the impact of 14 genetic variants on colorectal cancer phenotype and risk. *Am. J. Epidemiol.* 2012; 175: 1–10.

49. Marshall, KW, Mohr, S, Khettabi, F El, Nossova, N, Chao, S, Bao, W, et al. A blood-based biomarker panel for stratifying current risk for colorectal cancer. *Int. J. Cancer* 2010; 126: 1177–1186.
50. Taylor, DP, Stoddard, GJ, Burt, RW, Williams, MS, Mitchell, J a, Haug, PJ, et al. How well does family history predict who will get colorectal cancer? Implications for cancer screening and counseling. *Genet. Med.* 2011; 13: 385–391.
51. Wang, HM, Chang, TH, Lin, FM, Chao, TH, Huang, WC, Liang, C, et al. A new method for post Genome-Wide Association Study (GWAS) analysis of colorectal cancer in Taiwan. *Gene* 2013; 518: 107–113.
52. Wei, Y-S, Lu, J-C, Wang, L, Lan, P, Zhao, H-J, Pan, Z-Z, et al. Risk factors for sporadic colorectal cancer in southern Chinese. *World J. Gastroenterol.* 2009; 15: 2526–2530.
53. Wells, BJ, Kattan, MW, Cooper, GS, Jackson, L, Koroukian, S. ColoRectal Cancer Predicted Risk Online (CRC-PRO) Calculator Using Data from the Multi-Ethnic Cohort Study. *J. Am. Board Fam. Med.* 27: 42–55.
54. Yarnall, JM, Crouch, DJM, Lewis, CM. Incorporating non-genetic risk factors and behavioural modifications into risk prediction models for colorectal cancer. *Cancer Epidemiol.* 2013; 37: 324–9.
55. Amir, E, Freedman, OC, Seruga, B, Evans, DG. Assessing women at high risk of breast cancer: a review of risk assessment models. *J. Natl. Cancer Inst.* 2010; 102: 680–91.

56. Usher-Smith, JA, Emery, J, Kassianos, AP, Walter, FM. Risk Prediction Models for Melanoma: A Systematic Review. *Cancer Epidemiol. Biomarkers Prev.* 2014; 23: 1450–1463.
57. Jemal, A, Center, MM, DeSantis, C, Ward, EM. Global patterns of cancer incidence and mortality rates and trends. *Cancer Epidemiol. Biomarkers Prev.* 2010; 19: 1893–907.
58. Marchand, LL. Combined influence of genetic and dietary factors on colorectal cancer incidence in Japanese Americans. *J. Natl. Cancer Inst. Monogr.* 1999; 101–5.
59. Hewitson, P, Glasziou, P, Watson, E, Towler, B, Irwig, L. Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (hemocult): an update. *Am. J. Gastroenterol.* 2008; 103: 1541–9.
60. Scholefield, JH, Moss, SM, Mangham, CM, Whynes, DK, Hardcastle, JD. Nottingham trial of faecal occult blood testing for colorectal cancer: a 20-year follow-up. *Gut* 2012; 61: 1036–40.
61. Shaukat, A, Mongin, SJ, Geisser, MS, Lederle, FA, Bond, JH, Mandel, JS, et al. Long-term mortality after screening for colorectal cancer. *N. Engl. J. Med.* 2013; 369: 1106–14.
62. Atkin, WS, Edwards, R, Kralj-Hans, I, Wooldrage, K, Hart, AR, Northover, JMA, et al. Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial. *Lancet* 2010; 375: 1624–33.
63. Segnan, N, Armaroli, P, Bonelli, L, Risio, M, Sciallero, S, Zappa, M, et al. Once-only sigmoidoscopy in colorectal cancer screening: follow-up findings of the Italian Randomized Controlled Trial--SCORE. *J. Natl. Cancer Inst.* 2011; 103: 1310–22.

64. Brenner, H, Chang-Claude, J, Jansen, L, Knebel, P, Stock, C, Hoffmeister, M. Reduced risk of colorectal cancer up to 10 years after screening, surveillance, or diagnostic colonoscopy. *Gastroenterology* 2014; 146: 709–17.
65. Brenner, H, Stock, C, Hoffmeister, M. Effect of screening sigmoidoscopy and screening colonoscopy on colorectal cancer incidence and mortality: systematic review and meta-analysis of randomised controlled trials and observational studies. *BMJ* 2014; 348: g2467.
66. Nishihara, R, Wu, K, Lochhead, P, Morikawa, T, Liao, X, Qian, ZR, et al. Long-term colorectal-cancer incidence and mortality after lower endoscopy. *N. Engl. J. Med.* 2013; 369: 1095–105.
67. Lansdorp-Vogelaar, I, Knudsen, AB, Brenner, H. Cost-effectiveness of colorectal cancer screening. *Epidemiol. Rev.* 2011; 33: 88–100.
68. Pignone, M, Saha, S, Hoerger, T, Mandelblatt, J. Cost-effectiveness analyses of colorectal cancer screening: a systematic review for the U.S. Preventive Services Task Force. *Ann. Intern. Med.* 2002; 137: 96–104.
69. Provenzale, D. Cost-effectiveness of screening the average-risk population for colorectal cancer. *Gastrointest. Endosc. Clin. N. Am.* 2002; 12: 93–109.
70. Lansdorp-Vogelaar, I, Ballegooijen, M van, Zauber, AG, Boer, R, Wilschut, J, Winawer, SJ, et al. Individualizing colonoscopy screening by sex and race. *Gastrointest. Endosc.* 2009; 70: 96–108, 108.e1–24.

71. Seeff, LC, Manninen, DL, Dong, FB, Chattopadhyay, SK, Nadel, MR, Tangka, FKL, et al. Is there endoscopic capacity to provide colorectal cancer screening to the unscreened population in the United States? *Gastroenterology* 2004; 127: 1661–9.
72. Vijan, S, Inadomi, J, Hayward, RA, Hofer, TP, Fendrick, AM. Projections of demand and capacity for colonoscopy related to increasing rates of colorectal cancer screening in the United States. *Aliment. Pharmacol. Ther.* 2004; 20: 507–15.
73. Routine aspirin or nonsteroidal anti-inflammatory drugs for the primary prevention of colorectal cancer: U.S. Preventive Services Task Force recommendation statement. *Ann. Intern. Med.* 2007; 146: 361–4.
74. Dubé, C, Rostom, A, Lewin, G, Tsertsvadze, A, Barrowman, N, Code, C, et al. The use of aspirin for primary prevention of colorectal cancer: a systematic review prepared for the U.S. Preventive Services Task Force. *Ann. Intern. Med.* 2007; 146: 365–75.
75. Cuzick, J, Otto, F, Baron, JA, Brown, PH, Burn, J, Greenwald, P, et al. Aspirin and non-steroidal anti-inflammatory drugs for cancer prevention: an international consensus statement. *Lancet. Oncol.* 2009; 10: 501–7.
76. Joseph, DA, King, JB, Miller, JW, Richardson, LC. Prevalence of colorectal cancer screening among adults--Behavioral Risk Factor Surveillance System, United States, 2010. *MMWR. Morb. Mortal. Wkly. Rep.* 2012; 61 Suppl: 51–6.
77. Felsen, CB, Piasecki, A, Ferrante, JM, Ohman-Strickland, PA, Crabtree, BF. Colorectal cancer screening among primary care patients: does risk affect screening behavior? *J. Community Health* 2011; 36: 605–11.

78. Blalock, SJ, DeVellis, BM, Afifi, RA, Sandler, RS. Risk perceptions and participation in colorectal cancer screening. *Health Psychol.* 1990; 9: 792–806.
79. Beydoun, HA, Beydoun, MA. Predictors of colorectal cancer screening behaviors among average-risk older adults in the United States. *Cancer Causes Control* 2008; 19: 339–59.
80. Straus, WL, Mansley, EC, Gold, KF, Wang, Q, Reddy, P, Pashos, CL. Colorectal cancer screening attitudes and practices in the general population: a risk-adjusted survey. *J. Public Health Manag. Pract.* 11: 244–51.
81. Platz, EA, Willett, WC, Colditz, GA, Rimm, EB, Spiegelman, D, Giovannucci, E. Proportion of colon cancer risk that might be preventable in a cohort of middle-aged US men. *Cancer Causes Control* 2000; 11: 579–88.
82. Parkin, DM, Olsen, A-H, Sasieni, P. The potential for prevention of colorectal cancer in the UK. *Eur. J. Cancer Prev.* 2009; 18: 179–90.
83. Emmons, KM, McBride, CM, Puleo, E, Pollak, KI, Clipp, E, Kuntz, K, et al. Project PREVENT: a randomized trial to reduce multiple behavioral risk factors for colon cancer. *Cancer Epidemiol. Biomarkers Prev.* 2005; 14: 1453–9.
84. Houlston, RS. COGENT (COLOrectal cancer GENeTics) revisited. *Mutagenesis* 2012; 27: 143–51.

Table 1. Summary of 52 risk models.

Author, year	Country	Outcome	Factors included in score	Factors considered but not included	TRIPOD level*	Data source
Betes 2003a (21)	Spain	ACN +	Age, gender, BMI	NSAIDs, nonspecific abdominal pain, bowel habit (1-2 movements/day; diarrhoea-alternate; chronic constipation), cholesterol, triglycerides, form of recruitment	1a	Medical records
Betes 2003b (21)	Spain	ACN	Age, gender, BMI	NSAIDs, nonspecific abdominal pain, bowel habit (1-2 movements/day; diarrhoea-alternate; chronic constipation), cholesterol, triglycerides, form of recruitment	1a, 4	Medical records
Cai 2012 (22)	China	ACN	Age, gender, smoking, diabetes mellitus, green vegetables, pickled food, fried food, white meat	BMI, hypertension, hypertriglyceridaemia, alcohol intake, calcium or vitamin D supplementation, aspirin or NSAIDs, fresh fruit, eggs, milk, red meat	2a, 4	Questionnaire
Chen 2013 (23)	China	ACN	Age, smoking, alcohol	Gender, history of CVD, egg intake, defaecation frequency, education level, hypertension, diabetes, hyperlipidaemia, gastric / gallbladder / appendix operations history, aspirin, tea drinking, physical activity, green vegetable / fruit / milk / pickled food / fried or smoked food / bamboo root / red meat / white meat intake	1b	Medical records
Chen 2014 (24)	China	ACN	Age, gender, history of CHD, egg intake, defaecation frequency	Education level, hypertension, diabetes, hyperlipidaemia, gastric/ gallbladder / appendix operations history, aspirin use, smoking, alcohol, tea drinking, physical activity, green vegetable / fruit / milk / pickled food / fried or smoked food / bamboo root / red meat intake / white meat intake	1b	Questionnaire
Hassan 2013 (25)	Italy	ACN	Age, gender	Family history, BMI	1b	Medical records
Kaminski 2014 (26)	Poland	ACN	Age, gender, BMI, smoking, number and age affected of first degree relatives with CRC	Diabetes, regular aspirin use	2a	Questionnaire
Lin 2006 (27)	USA	ACN	Age, gender, first degree relative with CRC or second degree relative with adenoma	None	1a, 4	Questionnaire
Lin 2013 (28)	USA	ACN	Age, BMI, smoking, number of first degree relatives with CRC, previous sigmoidoscopy or colonoscopy, polyp history in past 10 years, physical activity, vegetable consumption, NSAID use, oestrogen use	None	1b	Questionnaire
Stegeman 2013 (30)	Netherlands	ACN	Age, gender, BMI, first degree relative with CRC, menopausal status (women), smoking, sleep, vigorous exercise, alcohol, fibre intake, calcium intake, red meat intake, aspirin/NSAID use	None	1a	Questionnaire
Stegeman 2014 (29)	Netherlands	ACN	Age, smoking, first degree relative with CRC, faecal immunochemical test, calcium intake	BMI, menopausal status, aspirin/NSAID use, fibre / red meat intake	1a	Questionnaire
Tao 2014a (31)	Germany	ACN	Age, gender, smoking, first-degree relative with CRC, alcohol, previous polyp, red meat consumption, NSAIDS, previous colonoscopy	BMI, physical activity, vegetable / fruit intake, HRT	3	Questionnaire
Yeoh 2011 (32)	Asia	ACN	Age, gender, smoking, first degree relative with CRC	Alcohol, diabetes	3	Questionnaire

Almurshed 2009 (33)	Saudi Arabia	CC	Region, marital status, education level, employment status, activity level, physical activity, knowledge of high-fibre diet	None	1a	Questionnaire
Camp 2002 (34)	USA	CC	Age, BMI, first degree relative with CRC, NSAID use, long term vigorous physical activity, Western diet, folic acid, calcium intake, lutein intake, refined grain intake, Prudent dietary pattern	Sex, hormone replacement therapy, smoking history, calorific intake, dietary fibre, total vegetable / fat intake, glycaemic index of intake, mutagen index, alcohol consumption	1a	Questionnaire
Colditz 2000 (35)	USA	CC	BMI, first degree relative with CRC, faecal occult blood test or sigmoidoscopy, aspirin, IBD, folate, vegetables, alcohol, height, physical activity, oestrogen replacement, fruits, fibre, saturated fat, smoking	None	1a, 4	Questionnaire
Driver 2007a (36)	USA	CC	Age, BMI, history of smoking	Weekly or daily alcohol use, intake of vegetables, intake of multivitamins, vitamin C, vitamin E, intake of cold cereal, physical activity, history of diabetes	1a	Medical records
Ma 2010a (38)	Japan	CC	Age, BMI, smoking, alcohol, physical activity	FH CRC, diabetes	3	Questionnaire
Wei E 2009 (40)	USA	CC	Age, BMI, smoking, current or past HRT, height, first degree relative with colon cancer, processed meat consumption, folate intake, physical activity, aspirin use, sigmoidoscopy or colonoscopy during follow up	None	1a	Questionnaire
Wei E 2004a (41)	USA	CC	Age, gender, BMI, smoking, alcohol, first degree relative with colon cancer, physical activity, height, processed meat, servings of beef, pork or lamb, folate intake, calcium intake	None	1a	Questionnaire
Shin 2014a (39)	Korea	CC (male)	Age, BMI, family history of cancer, height, fasting serum glucose, total serum cholesterol, alcohol, meat consumption	Smoking, exercise	3	Questionnaire and blood test
Shin 2014d (39)	Korea	CC (female)	Age, family history of cancer, height, fasting serum glucose, meat consumption	BMI, alcohol, smoking, exercise, female reproductive factors	3	Questionnaire and blood test
Freedman 2009b (37)	USA	Distal CC (male)	BMI, number of first degree relatives with CRC, prior negative sigmoidoscopy/colonoscopy, polyp history, aspirin and NSAID use	FOBT, multivitamin use, red meat / fruit / vegetable intake, alcohol intake, physical activity, smoking, age	1a, 4**	Questionnaire
Freedman 2009e (37)	USA	Distal CC (female)	Age, BMI, number of relatives with CRC, prior negative sigmoidoscopy/colonoscopy, polyp history, aspirin and NSAID use, oestrogen use in last 2 years	FOBT, multivitamin use, red meat / fruit / vegetable intake, alcohol intake, physical activity	1a, 4**	Questionnaire
Freedman 2009a (37)	USA	Proximal CC (male)	BMI, smoking, number of first degree relatives with CRC, prior negative sigmoidoscopy/colonoscopy, polyp history, aspirin and NSAID use, vegetable consumption	FOBT, multivitamin use, red meat / fruit intake, alcohol intake, physical activity, age	1a, 4**	Questionnaire
Freedman 2009d (37)	USA	Proximal CC (female)	Number of first degree relatives with CRC, prior negative sigmoidoscopy/colonoscopy, polyp history, physical activity, aspirin and NSAID use, vegetable consumption, oestrogen use in last 2 years	FOBT, multivitamin use, red meat / fruit intake, alcohol intake, BMI, age	1a, 4**	Questionnaire
Bener 2010 (42)	Qatar	CRC	BMI, smoking, family history of CRC, consumption of bakery products, consumption of soft drinks	Smoking of Sheesha, fresh fruit / fresh vegetable / green salad / frozen meat/chicken / fast food / processed food intake, consanguinity	1a	Questionnaire
Driver 2007b (36)	USA	CRC	Age, BMI, history of smoking, weekly or daily alcohol use	Intake of vegetables, intake of multivitamins, vitamin C, vitamin E, intake of cold cereal, physical activity, history of diabetes	1b	Medical records

Dunlop 2013 (43)	Worldwide	CRC	Age, gender, first degree relative with CRC, 10 SNPs ¹	None	3	Questionnaire and blood test for genetics
Guesmi 2010 (44)	Tunisia	CRC	Age, meat consumption, milk consumption	Gender, anaemia, smoking, physical activity, fruit / fried food intake, urban or rural living, olive oil consumption, walking	1a	Questionnaire
Han 2008 (45)	Not given	CRC	5 genes ²	Affymetrix U133Plus 2.0 chip	3	Blood test for genetics
Johnson 2013 (47)	Worldwide	CRC	BMI, smoking, first degree relative with CRC, physical activity, alcohol, IBD, hormone therapy (current or former), aspirin/NSAIDs, processed meat / red meat / fruit / vegetable intake	None	1a	Questionnaire
Lubbe 2012 (48)	UK	CRC	14 SNPs ³	None	1a	Blood test for genetics
Ma 2010c (38)	Japan	CRC	Age, BMI, smoking, physical activity, alcohol	FH CRC, diabetes	3	Questionnaire
Marshall 2010 (49)	Canada and USA	CRC	7 genes ⁴	38 genes	2b	Blood test for genetics
Tao 2014b (31)	Germany	CRC	Age, gender, smoking, first-degree relative with CRC, alcohol, previous polyp, red meat consumption, NSAIDs, previous colonoscopy	BMI, physical activity, vegetable / fruit intake, HRT	3	Questionnaire
Taylor 2011 (50)	USA	CRC	Age, first, second and third degree relatives with CRC	None	1a	Questionnaire
Wang 2013 (51)	Taiwan	CRC	16 SNPs ⁵	10 additional SNPs	1b	Blood test for genetics
Yarnall 2013 (54)	UK data	CRC	BMI, smoking, alcohol, fibre intake, red meat intake, physical activity, 14 SNPs ⁶	None	1a	Questionnaire and blood test for genetics
Wei Y 2009 (52)	China	CRC	BMI, smoking, first or second degree relative with CRC, alcohol	None	1a	Questionnaire
Shin 2014c (39)	Korea	CRC (male)	Age, BMI, family history of cancer, height, fasting serum glucose, total serum cholesterol, alcohol, meat consumption	Smoking, exercise	3	Questionnaire and blood test
Wells 2014b (53)	California and Hawaii	CRC (male)	Age, BMI, smoking, first degree relative with CC, race/ethnicity, alcohol, years of education, regular use of aspirin, multivitamins, red meat intake, history of diabetes, physical activity	History of cancer, regular use of NSAIDs, preference for well-done meat	1b	Questionnaire
Jo 2012b (46)	Korea	CRC (male)	3 SNPs ⁷ , age, family history of CRC	From 426,019 SNPs	1b	Questionnaire and blood test for genetics
Shin 2014f (39)	Korea	CRC (female)	Age, family history of cancer, height, fasting serum glucose, meat consumption	BMI, alcohol, smoking, exercise, female reproductive factors	3	Questionnaire and blood test
Wells 2014a (53)	California and Hawaii	CRC (female)	Age, BMI, smoking, first degree relative with CC, race/ethnicity, alcohol, years of education, regular use of NSAIDs, multivitamins, history of diabetes, use of oestrogen	Preference for well done meat, physical activity, regular use of aspirin, red meat intake, history of cancer	1b	Questionnaire
Jo 2012a	Korea	CRC (female)	Age, family history of CRC, 5 SNPs ⁸	From 426,019 SNPs	1b	Questionnaire

(46)						and blood test for genetics
Ma 2010b (38)	Japan	Rectal cancer	Age, BMI, physical activity, alcohol	FH CRC, diabetes, smoking	3	Questionnaire
Wei E 2004b (41)	USA	Rectal cancer	Age, BMI, smoking, first degree relative with rectal cancer, alcohol, physical activity, height, processed meat, gender, servings of beef, pork or lamb, folate intake, calcium intake	None	1a	Questionnaire
Freedman 2009c (37)	USA	Rectal cancer (male)	Number of first degree relatives with CRC, prior negative sigmoidoscopy/colonoscopy, polyp history, NSAID use, physical activity	FOBT, multivitamin use, red meat / fruit / vegetable intake, alcohol intake, smoking, BMI, age	1a, 4*	Questionnaire
Shin 2014b (39)	Korea	Rectal cancer (male)	Age, BMI, family history of cancer, height, fasting serum glucose, total serum cholesterol, alcohol, meat consumption	Smoking, exercise	3	Questionnaire and blood test
Freedman 2009f (37)	USA	Rectal cancer (female)	BMI, number of first degree relatives with CRC, prior negative sigmoidoscopy/colonoscopy, polyp history, physical activity, NSAID use, oestrogen use in last 2 years	FOBT, multivitamin use, red meat / fruit / vegetable intake, alcohol intake, age	1a, 4*	Questionnaire
Shin 2014e (39)	Korea	Rectal cancer (female)	Age, family history of cancer, height, fasting serum glucose, meat consumption	BMI, alcohol, smoking, exercise, female reproductive factors	3	Questionnaire and blood test

* Types of prediction model studies for each model defined according to the TRIPOD guidelines. 1a – Development only; 1b – Development and validation using resampling; 2a – Random split-sample development and validation; 2b – Nonrandom split-sample development and validation; 3 – Development and validation using separate data; 4 – Validation study

** The validation was for colon and rectal cancer combined

ACN+ - advanced colorectal neoplasia including moderate dysplasia

ACN – advanced colorectal neoplasia

CC – colon cancer

CRC – colorectal cancer

BMI – body mass index

NSAIDs – non-steroidal anti-inflammatory drugs

CVD – cardiovascular disease

CHD – coronary heart disease

HRT – hormone replacement therapy

IBD – inflammatory bowel disease

FOBT – faecal occult blood test

FH – family history

¹ rs6983267, rs4779584, rs4939827, rs3802842, rs10795668, rs16892766, rs4444235, rs9929218, rs10411210, rs961253

² BANK1, B-cell scaffold protein with ankyrin repeats 1; BCNP1, B-cell novel protein 1; CDA, cytidine deaminase; MGC20553, FERM domain containing 3; MS4A, membrane-spanning 4 domains

³ 14 SNPs localizing to 14 chromosome regions – 1q41, 3q26.2, 8q23.3, 8q24.21, 10p14, 11q23.1, 12q13.13, 14q22.2, 15q13.3, 16q22.1, 18q21.1, 19q13.11, 20p12.3, 20q13.33

⁴ ANXA3, Annexin A3; CLEC4D, C-type lectin domain family4, member D; IL2RB, Interleukin 2 receptor, beta; LMNB1, Lamin B1; PRRG4, Proline risk Gla 4; TNFAIP6, Tumour necrosis factor, alpha-induced protein 6; VNN1, Vanin 1

⁵ rs1983891, rs869736, rs3214050, rs10411210, rs3731055, rs231775, rs1412829, rs1572072, rs6983267, rs1799782, rs712221, rs160277, rs11721827, rs2736100, rs3135967, rs1760944

⁶ rs6691170, rs10936599, rs16892766, rs6983267, rs10795668, rs3802842, rs11169552, rs4444235, rs4779584, rs9929218, rs4939827, rs10411210, rs961253, rs4925386

⁷ rs17391002, rs9549448, rs254833

⁸ rs10083736, rs16987827, rs8046516, rs9926182, rs175237

Table 2. 87 Risk factors (excluding genes and SNPs) considered across all included studies

<p>Personal characteristics Age BMI Gender Consanguinity Family history of colorectal cancer Height Race / ethnicity Marital status Education level Employment status Knowledge of high-fibre diet Years of education Urban or rural living</p> <p>Personal medical history Gastric operation history Gallbladder operation history Appendix operations Hypertension Diabetes or history of diabetes Inflammatory bowel disease History of coronary heart disease History of cardiovascular disease Polyp history History of cancer Defaecation frequency Non-specific abdominal pain</p> <p>Female hormonal factors HRT (ever, current or past) Oestrogen use Menopausal status Age at menarche Age at first childbirth Age at menopause</p> <p>Lifestyle Smoking (tobacco or Sheesha) Alcohol Physical activity Sleep</p> <p>Drug and vitamin supplementation NSAID use Aspirin use Multivitamin use Calcium supplementation Vitamin D supplementation Vitamin C supplementation Vitamin E supplementation</p>	<p>Diet Fibre intake Meat Red meat Processed meat Servings of beef, pork or lamb White meat Frozen meat/chicken Preference for well-done meat Vegetables Fresh vegetables Green vegetables Green salad Fruit Fast food Processed food Pickled food Fried food Smoked food Eggs Milk Fat Saturated fat Bakery products Refined grain Tea Olive oil Soft drinks Bamboo root intake Cold cereal Glycaemic index of intake Western diet Prudent dietary pattern Calorific intake Mutagen index* Calcium intake Folic acid intake Lutein intake</p> <p>Biomarkers Fasting glucose Hyperlipidaemia Cholesterol Triglycerides Haemoglobin</p> <p>Other tests Faecal immunochemical test Faecal occult blood test Prior sigmoidoscopy or colonoscopy</p>
---	--

Table 3. Details of performance of models

Author, year	Outcome	Performance in development population			Performance in bootstrap or cross-validation			Performance in sub-set of population			Performance in external population			
		Discrimination AUROC (95% CI)	Calibration	Accuracy	Discrimination AUROC (95% CI)	Calibration	Accuracy	Discrimination AUROC (95% CI)	Calibration	Accuracy	Discrimination AUROC (95% CI)	Calibration	Accuracy	External reference
Betes 2003a (21)	ACN+	0.65		PPV 12.0-50.0										
Betes 2003b (21)	ACN	0.67		PPV 7.3-33.3										Cai 2012 (22)
Cai 2012 (22)	ACN	0.74 (0.72-0.77)		Sens 82.8; Spec 50.8	0.74 (0.72-0.77)			0.74 (0.70-0.78)	H-L p=0.77	Sens 80.3; Spec 51.2	0.65 (0.64-0.78)			Chen 2014 (24)
Chen 2013 (23)	ACN	0.65 (0.61-0.69)	H-L p=0.093	Sens 65.1; Spec 57.2; PPV 44.4; NPV 75.7	0.66 (0.62-0.68)									Chen 2014 (24)
Chen 2014 (24)	ACN	0.75 (0.69-0.82)	H-L p=0.205	Sens 93.8; Spec 47.6; PPV 9.1; NPV 99.3	0.75 (0.70-0.82)									
Hassan 2013 (25)	ACN					H-L p=0.30								
Kaminski 2014 (26)	ACN	0.64*	H-L p=0.74*					0.62 (0.60-0.64)	E/O ratio 1 (0.95-1.06). H-L p=0.16*	Sens 92.4, Spec 13.9, PPV 7.55, NPV 96.0;				
Lin 2006 (27)	ACN										0.65 (0.61-0.70)			Cai 2012 (22)
Lin 2013 (28)	ACN				Men 0.61 (0.58-0.65), Women 0.62 (0.58-0.66)						0.71 (0.64-0.77)			Chen 2014 (24)
Stegeman 2014 (29)	ACN	0.76	H-L p=0.94	Sens 40; Spec 93										
Tao 2014a (31)	ACN	0.67 (0.65-0.69)	H-L p=0.21								0.66 (0.63-0.69)	H-L p=0.65		Tao 2014 (31)
Yeoh 2011 (32)	ACN	0.66 (0.62-0.70)	H-L p=0.29								0.64 (0.60-0.68)	H-L p=0.49		

Colditz 2000 (34)	CC					Women 0.67 (0.64-0.70); Men 0.71 (0.68-0.74)**	Kim 2004 (17)
						0.6	Schroy 2012 (19)
Driver 2007a (36)	CC	0.72	H-L p=0.43				
Ma 2010a (38)	CC	0.71 (0.68-0.74)				0.66(0.62-0.70)	χ^2 p=0.20; E/O 1.19 (1.03-1.37) Ma 2010 (38)
Wei E 2009 (40)	CC	0.61 (0.59-0.63)					
Shin 2014a (39)	CC (male)	0.77 (0.76-0.78)	χ^2 p=0.22			0.77 (0.75-0.79)	χ^2 p=0.029
Shin 2014d (39)	CC (female)	0.71 (0.69-0.73)	χ^2 p=0.73			0.72 (0.70-0.74)	χ^2 p=0.49
Driver 2007b (36)	CRC	0.70	H-L p=0.91	0.69			
Dunlop 2013 (43)	CRC			0.59	PPV 0.71; NPV 0.51	0.57	Dunlop 2013 (43)
Han 2008 (45)	CRC	0.88 (0.81-0.94)	Sens 94; Spec 77 PPV 82, NPV 92		79% (71.5-86.5)		Sens 88; Spec 64. PPV 67; NPV 87
Ma 2010c (38)	CRC	0.70 (0.68-0.72)				0.64 (0.61-0.67)	χ^2 p=0.08; E/O 1.09 (0.98-1.23) Ma 2010 (38)
Marshall 2010 (49)	CRC	0.80 (0.74-0.85)	Sens 82; Spec 64; PPV 68, NPV 79		0.80 (0.76-0.84)	Sens 72; Spec 70; PPV 70, NPV 72	Sens 71.7; Spec 71.2 Yip 2010 (20)
Tao 2014b (31)	CRC	0.71 (0.67-0.75)				0.68 (0.57-0.79)	Tao 2014 (31)
Taylor 2011 (50)	CRC	0.67					
Wang 2013 (51)	CRC	0.77		0.72			
Yarnall 2013 (54)	CRC	0.63					

Freedman 2009a,b,c (37)	CRC (male)				0.61 (0.60-0.62)	E/O ratio 0.99 (0.96-1.04)	Park 2008 (18)
Jo 2012b (46)	CRC (male)	0.73 (0.68-0.77)		0.70 (0.65-0.74)			
Shin 2014c (39)	CRC (male)	0.76 (0.76-0.77)	χ^2 p=0.1035		0.78 (0.77-0.79)	χ^2 p=0.0003	
Wells 2014b (53)	CRC (male)	0.69		0.68 (0.67-0.69)			
Freedman 2009d,e,f (37)	CRC (female)				0.61 (0.59-0.62)	E/O ratio 1.05 (0.98-1.11)	Park 2008 (18)
Jo 2012a (46)	CRC (female)	0.65 (0.62-0.68)		0.60 (0.56-0.64)			
Shin 2014f (39)	CRC (female)	0.71 (0.70-0.72)	χ^2 p=0.6123		0.73 (0.71-0.74)	χ^2 p=0.1569	
Wells 2014a (53)	CRC (female)	0.69		0.68 (0.67-0.69)			
Ma 2010b (38)	Rectal cancer	0.68 (0.64-0.71)			0.62 (0.57-0.66)	χ^2 p=0.19; E/O 0.94 (0.78-1.12)	Ma 2010 (38)
Shin 2014b (39)	Rectal cancer (male)	0.75 (0.74-0.76)	χ^2 p=0.29		0.78 (0.77-0.79)	χ^2 p=0.0003	
Shin 2014e (39)	Rectal cancer (female)	0.70 (0.68-0.71)	χ^2 p=0.084		0.72 (0.70-0.74)	χ^2 p=0.198	

ACN+ Advanced colorectal neoplasia including moderate dysplasia

ACN – advanced colorectal neoplasia defined as invasive cancer, an adenoma 10mm or more, a villous adenoma (at least 25% villous) or an adenoma with high grade dysplasia

CC – colon cancer

CRC – colorectal cancer

AUROC – area under the receiver operator curve. Values given as mean and 95% confidence intervals

Sens – sensitivity

Spec – specificity

PPV – positive predictive value

NPV – negative predictive value

H-L – Hosmer-Lemeshow test

E/O – Expected over observed ratio

* These values are from the model prior to conversion of the coefficients to scores (Kaminski)

** Removed aspirin use from men and history of chronic IBD from both genders as not available so actually not validating original score

FIGURE LEGENDS

Figure 1. PRISMA flow diagram

Figure 2. Relative discriminative performance of the risk scores ordered by number and complexity of variables included

Figure 1: PRISMA Flow Diagram

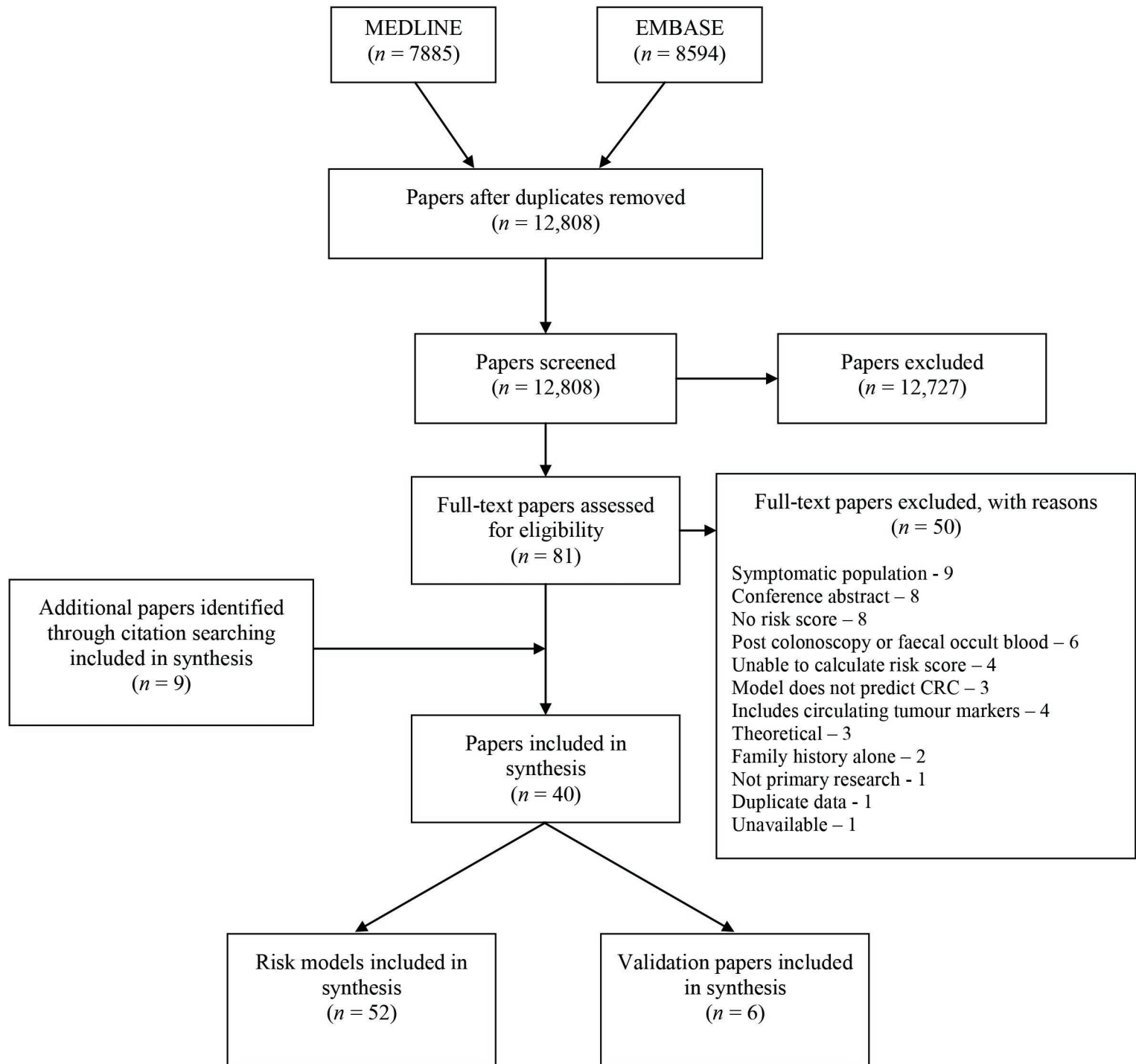


Figure 2

