

Systematic Review

From Algorithms to Clinical Utility: A Systematic Review of Individualized Risk Prediction Models for Colorectal Cancer

Deborah Jael Herrera ¹, Wessel van de Veerdonk ^{1,2}, Daiane Maria Seibert ³, Moges Muluneh Boke ¹,
Claudia Gutiérrez-Ortiz ¹, Nigus Bililign Yimer ¹, Karen Feyen ³, Allegra Ferrari ^{1,4} and Guido Van Hal ^{1,*}

- ¹ Family Medicine and Population Health Department (FAMPOP), Faculty of Medicine and Health Sciences, University of Antwerp, 2610 Antwerp, Belgium; djherrera5947@gmail.com (D.J.H.); wessel.vandeveerdonk@thomasmore.be (W.v.d.V.); mogelove75@gmail.com (M.M.B.); claudia.gutierrezortiz@student.uantwerpen.be (C.G.-O.); nigusb27@gmail.com (N.B.Y.); allegra.ferrari@uantwerpen.be (A.F.)
- ² Centre of Expertise Care and Well-Being, Campus Zandpoortvest, Thomas More University of Applied Sciences, 2800 Mechelen, Belgium
- ³ Centre of Expertise Design and Technology, Campus De Nayer, Thomas More University of Applied Sciences, 2800 Mechelen, Belgium; daiane.seibert@thomasmore.be (D.M.S.); karen.feyen@thomasmore.be (K.F.)
- ⁴ Department of Health Sciences (DISSAL), University of Genoa, Via Pastore 1, 16123 Genoa, Italy
- * Correspondence: guido.vanhal@uantwerpen.be

Abstract: Individualized risk prediction models for colorectal cancer (CRC) play a pivotal role in shaping risk-based screening approaches, garnering attention for use in informed decision making by patients and clinicians. While the incorporation of new predictors and the development of advanced yet complex prediction models can enhance model performance, their practical implementation in clinical settings remains challenging. This systematic review assessed individualized CRC risk prediction models for their validity and potential clinical utility. Utilizing the Cochrane Collaboration methods and PROBAST tool, we conducted comprehensive searches across key databases and risk of bias assessment, respectively. Out of 41 studies included evaluating 44 risk prediction models, 12 conventional and 3 composite models underwent external validation. All risk models exhibited varying discriminatory accuracy, with the area under the curve (AUCs) ranging from 0.57 to 0.90. However, most studies showed an unclear or high risk of bias, with concerns about applicability. Of the five models with promising clinical utility, only two underwent external validation and one employed a decision curve analysis. These models demonstrated a discriminating and well-calibrated performance. While high-performing CRC risk prediction models exist, a need for transparent reporting of performance metrics and their clinical utility persists. Further research on this area is needed to facilitate the integration of these models into clinical practice, particularly in CRC screening.

Keywords: risk prediction; risk scores; colorectal cancer; advanced neoplasia; risk factors; model performance; clinical utility



Citation: Herrera, D.J.; van de Veerdonk, W.; Seibert, D.M.; Boke, M.M.; Gutiérrez-Ortiz, C.; Yimer, N.B.; Feyen, K.; Ferrari, A.; Van Hal, G. From Algorithms to Clinical Utility: A Systematic Review of Individualized Risk Prediction Models for Colorectal Cancer. *Gastrointest. Disord.* **2023**, *5*, 549–579. <https://doi.org/10.3390/gidisord5040045>

Academic Editor: Tamara Čačev

Received: 18 September 2023

Revised: 20 October 2023

Accepted: 7 November 2023

Published: 11 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Colorectal cancer (CRC) is the third most frequently diagnosed cancer and the second leading cause of cancer death worldwide for both sexes combined [1]. The past two decades have witnessed a notable decline in CRC incidences and mortality, attributed in part to population-based screening, precursor lesion removal, and shifts toward healthier population-wide lifestyles, particularly in Western countries [2–5]. However, despite the substantial progress made in CRC prevention through these screening efforts, they remain insufficient in providing personalized prognostic information for individual patients [1].

Traditionally, cancer screening has been directed by population-wide guidelines that advocate for uniform screening approaches, often limited to age and, in certain cases, a family history of cancer, without further differentiation [6,7]. While these guidelines have

undoubtedly contributed to the early detection of many cancers, they may inadvertently overlook the heterogeneity that exists within these groups. Individuals possess unique genetic makeups, lifestyle choices, and clinical profiles that significantly influence their risk of developing cancer [8]. Failing to account for these important traits can result in both over-screening for some and under-screening for others, thereby diluting the impact of screening efforts [9]. The complex interplay between these multifaceted risk factors and the broader public health strategies necessitates a more nuanced and personalized approach for decision making regarding prevention and screening [10]. On the other hand, in countries where increasing CRC incidence is experienced without organized screening programs, it is crucial to explore alternative approaches that can effectively identify and prioritize high-risk individuals for screening colonoscopy [11]. With this regard, individualized risk prediction models have emerged as a pivotal tool that can be combined with other (medical) decision-making tools to support informed choice and enhance personalized care in CRC screening [12].

An individualized risk prediction model is defined as a mathematical function that is designed to estimate the personal probability or (absolute) risk of developing a specific condition in the future based on two or more risk factors [13]. In the context of CRC, prediction models incorporate well-established risk factors and sometimes use risk scoring systems and risk assessment tools or calculators to estimate the individual's likelihood of developing CRC or advance neoplasia (AN) [5,8,14]. There are generally two known types of risk models for CRC: conventional risk models and composite (advanced) risk models. Conventional risk predictors refer to the inclusion of readily available risk factors, such as age, sex, demographics, family history of the disease, and well-established comorbidities [15]. However, a growing interest on the inclusion of new and more advanced features to enhance the performance and effectiveness of risk assessment for CRC or AN has led to the development of composite models [16]. These models refer to those tools that leverage a combination of risk factors, clinical variables (e.g., laboratory/test results), and sometimes biomarkers or genetic information to estimate an individual's risk of developing CRC.

Primary care providers and patients both recognized the usefulness of risk prediction tools in supporting informed choice and its potential for enhancing participation, safety, and efficiency in CRC screening [12,17]. However, several barriers have offset this potential, including skepticism about the tool's accuracy, the complexity in assessing the predictors, and the categorizations of different risk groups that are not in line with the established screening guidelines [17–19]. To address these concerns, it is important to consider four key factors, including the performance, complexity (or clinical usability), potential clinical utility, and generalizability of the models, which determine their readiness for integration into risk-based CRC screening [10,20,21]. While sophisticated prediction models might promise heightened predictive accuracy, their practical application within clinical settings can be impeded by challenges due to their complexity, low potential for clinical utility, and limited generalizability [18,21]. Model complexity refers to the level of difficulty associated with implementing and incorporating a model into clinical practice, and it is assessed through an evaluation of the model's usability [22]. This evaluation considers whether the model utilizes straightforward risk scores, calculators, or nomograms to estimate the risk of CRC and facilitate efficient risk-stratification approaches. Conversely, the potential clinical utility of a model is determined by its interpretability of the performance metrics to clinicians and patients and its practicality in clinical settings, ensuring accessibility to both clinicians and patients for informed decision making [22]. The generalizability of the model, on the other hand, refers to whether the model has undergone external validation using appropriate exclusion criteria and a statistical method that accounts for confounding or competing risk, indicating its validity beyond its development context.

Although several systematic reviews are available related to CRC risk prediction models, these reviews solely focused on evaluating conventional risk scores, leaving a gap in understanding the potential benefits and trade-offs of incorporating more complex predictors for improved CRC and AN risk prediction [8,23]. Furthermore, these reviews

did not adequately consider the models' generalizability and did not assess the clinical usability and potential clinical utility of these risk scores. At its core, this study aimed to synthesize and appraise CRC risk prediction models to identify plausible models that are suited for integration into clinical practice. This involved a comprehensive examination of the model's performance, complexity, potential clinical utility, and generalizability in the context of predictive models for real-world clinical practice, particularly in CRC screening. Moreover, this review provides practical guidance for researchers to consider common pitfalls during model development, validation, and reporting of CRC risk prediction models that go beyond model performance.

2. Methods

As part of the ORIENT Project (Towards Informed Decisions in Colorectal Cancer Screening in Flanders) (<https://thomasmore.be/nl/orient> (accessed on 20 October 2023)), this review aims to synthesize evidence on the individualized risk prediction models for CRC or AN. Specifically, our review addresses the following key questions: What is the current state of evidence regarding the performance and validation of risk prediction models for CRC or AN? Which of these models have demonstrated readiness for adoption in CRC screening and support informed decision making in various CRC screening contexts?

The protocol of this systematic review was registered with the International Prospective Register of Systematic Reviews (PROSPERO registration number: CRD42022368227, https://www.crd.york.ac.uk/prospero/export_details_pdf.php (accessed on 13 September 2023)). Furthermore, this review adhered to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines for transparent and comprehensive reporting (Supplementary Materials Supplementary S1).

2.1. Criteria for Considering Studies for This Review

In this review, the Population, Index model, Comparator, Outcome, Timing, and Setting (PICOTS) system was employed to formulate the review question and guide the systematic review of risk prediction modeling studies [24] (Table 1).

Table 1. The PICOTS system for this review.

P	Population	Apparently healthy individuals at time of predictor assessment who had not undergone CRC screening (colonoscopy) and had no history of a CRC diagnosis or treatment within the past 10 years at the time of prediction/predictor assessment. Studies involving participants with Lynch syndrome or who have been diagnosed with CRC for the past 10 years were excluded.
I	Index model(s)	All risk prediction models (with or without external validation) that aimed to estimate the risk of developing CRC or AN; risk models that comprised individualized risk factors, including conventional and composite models. Studies were excluded if predictors were based solely on genetic information, test/laboratory findings, or a combination of both only.
C	Comparator	No predefined comparator.
O	Outcome(s)	CRC or AN detected during screening colonoscopy and cumulative risk of CRC.
T	Timing	The moment of predictor assessment before the screening for CRC; any moment before diagnosis was included.
S	Setting	Not specified.

2.2. Search Methods for Identification of Studies

Electronic Searches

The following databases were searched for English published studies from 1 January 2010 to 24 November 2022 and was updated on 3 October 2023: (1) Cochrane CENTRAL via Ovid, (2) MEDLINE via Ovid, (3) PubMed, and (4) Web of Science. For ongoing trials, the following database was searched: ClinicalTrials.gov (www.clinicaltrials.gov, accessed on 24 November 2022). To mitigate publication bias, additional sources were utilized, including searching references in relevant reviews of the literature or snowballing, and searching for grey literatures.

2.3. Selection of Studies

Articles retrieved from the electronic databases were exported directly as EndNote X9 files and imported into Rayyan software for de-duplication and screening (<https://www.rayyan.ai> (accessed on 9 October 2023)). During the initial screening, independent review authors (DJH, WvdV, AF, and DMS) assessed titles and abstracts, categorizing studies as “included”, “excluded”, or “maybe” for full-text screening based on the review’s inclusion criteria. Disagreements were resolved through consensus or arbitration by other review authors (GVH). In the subsequent round, full-text papers were obtained from the selected abstracts, allowing for a detailed assessment, including the identification of fraudulent or flawed studies and assessment of risk of bias. Regular group discussions were held among the members of the review team to ensure congruency and maintain written records.

2.4. Data Extraction and Data Management

Data extraction with a validation step was implemented to ensure the accuracy and consistency of the collected information. This process involved an independent review by the authors (DJH, DMS, KF, MHB, CGO, and NBY), who critically extracted, reviewed, and cross-verified the extracted data. Throughout this procedural sequence, ongoing dialogues among the review authors played a pivotal role in maintaining the consistency of the extracted data. Any disparities in interpretations or findings were effectively addressed and resolved through collaborative consensus. In instances where a unanimous agreement could not be reached, arbitration was proficiently overseen by the third set of review authors (GVH and WvdV).

2.5. Risk of Bias and Applicability Assessment

The RoB and applicability of risk prediction models from eligible studies was assessed using the PROBAST tool, following recommendations from the Cochrane Prognosis Methods Group [25,26]. The PROBAST tool enables a focused and transparent approach to assess the RoB and applicability of studies that develop, validate, or update prediction models for individualized predictions [27] (Figure 1). It consists of four domains: patient selection, predictors, outcome, and analysis. Each domain is assigned a risk of bias category: “low risk,” “high risk,” or “unclear risk.” To guide the assessment, signaling questions were employed with response options, such as “yes,” “probably yes,” “probably no,” and “no.” Domain-level judgements were made based on criteria fulfillment and available information [26,28]. This included the first three domains (patient selection, predictors, and outcome), which were also assessed for applicability concerns. Overall risk of bias was determined by combining domain-level judgments. Additionally, RoB and applicability assessments were conducted by six independent review authors (DJH, DMS, KF, MHB, CGO, and NBY). Consistency in judgement was maintained through discussions, disagreements were resolved through consensus, and arbitration by a third review author (DJH, DMS, and KF) was available if necessary.

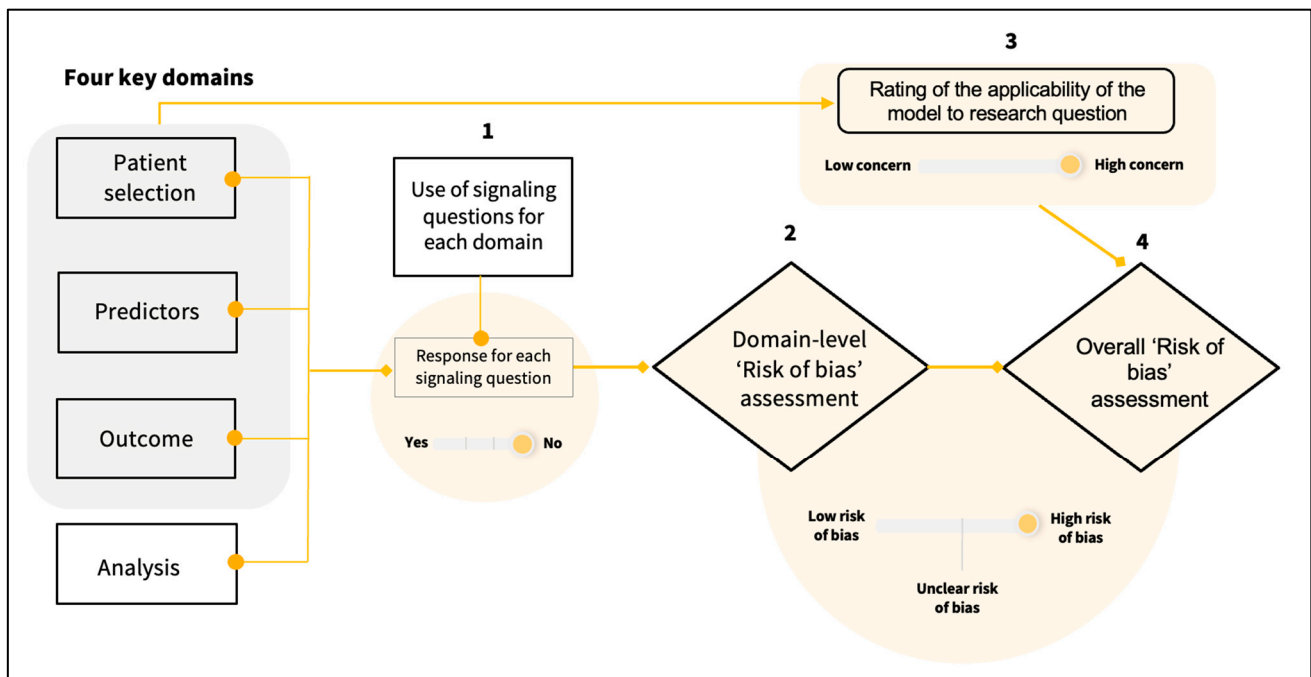


Figure 1. Schematic diagram of the “risk of bias” assessment based on the PROBAST tool [27].

2.6. Measures of Prediction Model Performance

Predictive performance measures were extracted from development and validation studies by using the following statistical measures: discriminatory accuracy based on area under the curve (AUC) values with 95% CIs, if available, and model calibration, including E/O ratios and the Hosmer–Lemeshow goodness-of-fit test (PHL) [13,29]. An AUC of 0.5 indicated a chance-level performance, and 1 reflected a perfectly discriminating model. Moreover, in cases where the E/O ratio was not reported, we estimated model performance using the following available formulas and reported quantities:

$$\ln(O : E) = \ln(O) - \ln(E)$$

and

$$SE(\ln(O : E)) = \frac{\sqrt{[N \times (\frac{O}{N}) \times (1 - \frac{O}{N})]}}{O}$$

where *O* is the total number of observed events, *E* is the total number of expected events, *N* is the total sample size, and *SE* is the standard error. Other metrics were further noted, including the range of sensitivity and specificity of the models, and net reclassification improvement (NRI), both relevant for assessing the model performance and potential clinical utility (Table 2).

Table 2. Definition of metrics used in evaluating model performance and potential clinical utility (adapted and modified from Wynants et al. 2019) [30].

Terms	Definition
AUC	In this case, the receiver operating characteristic curve. A measure of discrimination. For prediction models based on logistic regression; this corresponds to the probability that a randomly selected diseased patient had a higher risk prediction than a randomly selected patient who does not have the disease.
Calibration	Correspondence between predicted and observed risks is usually assessed in calibration plots or by calibration intercepts and slopes.

Table 2. *Cont.* [30].

Terms	Definition
Sensitivity	The proportion of true positives in truly diseased patients
Specificity	The proportion of true negatives in truly non-diseased patients.
Positive predictive value	The proportion of true positives in patients classified as positive.
Negative predictive value	The proportion of true negatives in patients classified as negative.
Decision curve analysis	A method to evaluate classifications for a range of possible thresholds, reflecting different costs of false positives and benefits of true positives.
Net reclassification improvement	Net reclassification improvement, reflecting reclassifications in the right direction when making decisions based on one prediction model compared to another

2.7. Dealing with Lack of Information in Included Studies

In terms of missing information or incomplete data on model performance, specifically on 95% CIs and E/O ratios, these measures were estimated using the formula proposed by Debray and colleagues [13,29].

2.8. Generalizability, Clinical Utility, and Usability of the Models

We undertook an assessment of the model's generalizability, potential clinical utility, and usability to identify models that are suitable for adoption in clinical practice. Our assessment of generalizability revolved around a model's external validation either in the original work or subsequent research. We critically assessed any potential selection bias and scrutinized for unwarranted participant exclusions, as well as the adequate sample size used.

In terms of clinical utility, our criteria focused on metrics that hold practical value in clinical settings, ensuring accessibility to both clinicians and patients for informed decision making. According to these criteria, we expected studies to report and discuss essential metrics, such as sensitivity, specificity, NPV, and PPV, as well as the number needed to screen (NNS) or refer (NRR) (Table 2). Additionally, we evaluated whether studies reported specific clinical scenarios or decision nodes, like net benefits and decision curves, associated with risk categories or multiple plausible thresholds. Notably, our evaluation was also centered on the appropriateness of determining risk thresholds to assess the model's potential clinical utility. Furthermore, in terms of usability, we took note of whether authors included tools such as calculators, nomograms, or reported absolute predicted probability estimates, along with simple risk scoring systems, all aimed at facilitating more accessible knowledge translation [22].

2.9. Data Synthesis and Hierarchical Clustering Approaches

Performance measures were extracted and summarized from all included studies, including AUC values, O/E ratios, and Hosmer–Lemeshow goodness-of-fit test [31–33]. Key study characteristics and the model performance of the models were presented in tables, with AUCs being further illustrated using forest plots. Additionally, we performed hierarchical clustering analysis using R software version 4.2.2 [34], visualized through a radial bar chart to provide a broad overview of the included models based on five domains: sociodemographics, behavior, clinical factors, patient history, laboratory/test findings limited to fecal immunochemical test (FIT)/fecal hemoglobin concentration (Fhbc) tests, and biomarkers. The aim was to explore patterns, specifically the similarities and the differences between prediction models within clusters and across different studies. On the other hand, an evidence matrix was used to visualize and assess the clinical usability of the models, as shown in the Supplementary Materials Figure S1. Finally, the RoB and applicability assessment was presented in detail, including the judgement for each domain, the reasons for such judgement, and the overall judgement, using a table following the PROBAST guideline (Supplementary Materials Supplementary S3).

3. Results

3.1. Study Inclusion

The initial electronic search generated a substantial dataset of 4852 records, including 1658 citations from MEDLINE via Ovid, 1802 citations from PubMed, 175 citations from Cochrane Central, and 1217 citations from Web of Science (Figure 2). After the removal of duplicates ($n = 1611$) and initial screening of studies based on their title and abstract ($n = 3806$), a total of 144 records qualified for full-text screening. Additionally, 19 studies were identified by citation mining of three literature reviews relevant to risk prediction models for CRC [8,23,35]. During the review update from November 2022 to October 2023, we identified five new eligible studies. Following full-text screening, 41 studies were considered for inclusion in this review.

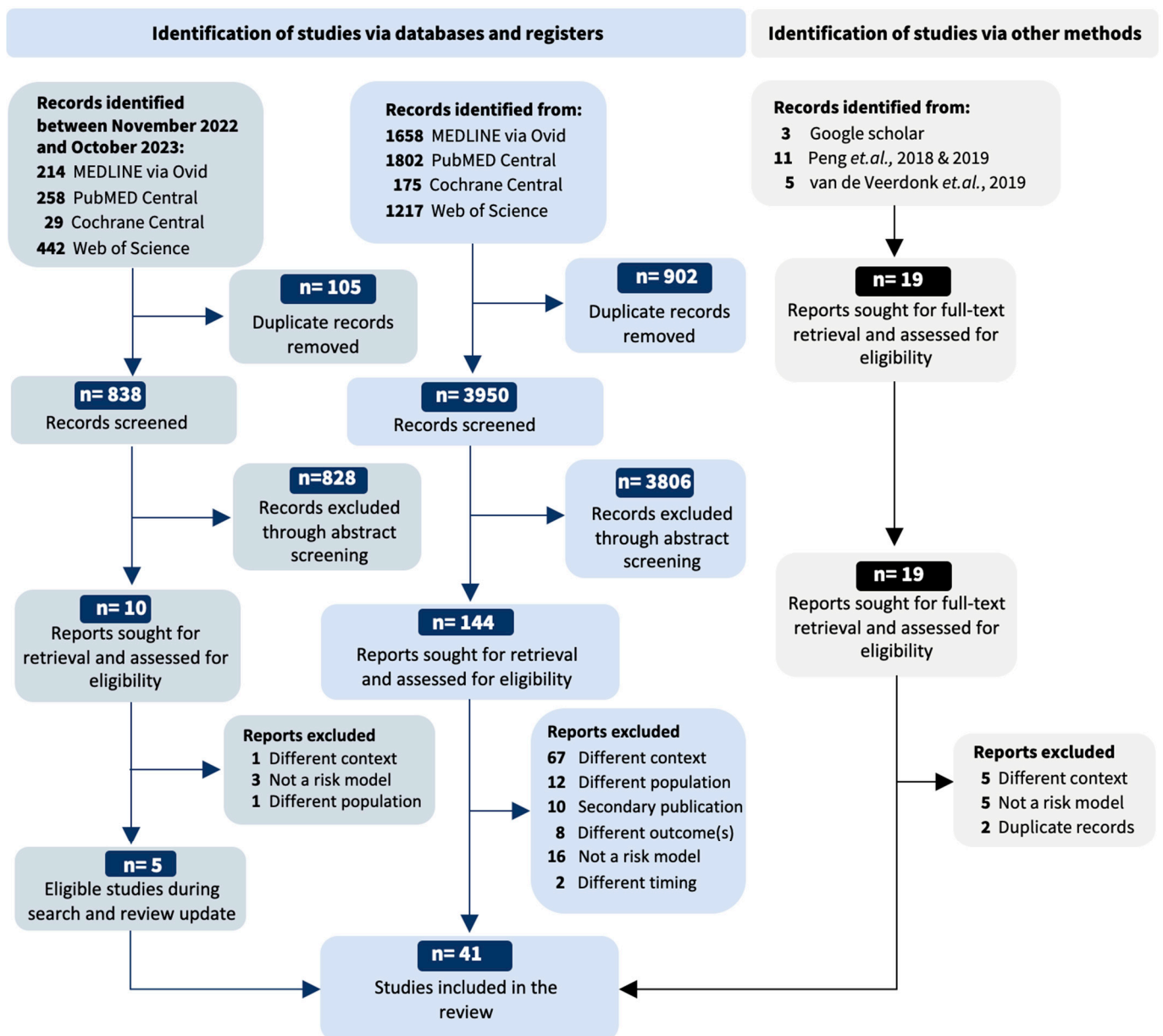


Figure 2. PRISMA flowchart [8,23,35].

3.2. Risk of Bias and Applicability of the Included Studies

Among the 41 studies appraised, 10 showed an overall unclear risk of bias [36–44], while 30 exhibited an overall high risk of bias [45–75]. Interestingly, only one study emerged with

an overall low risk of bias [76] (Figure 3). The primary sources of bias were traced to issues like the inappropriate exclusion of relevant participants, such as those with a family history of CRC and comorbidities like diabetes or cardiovascular diseases, ultimately diminishing the generalizability of the model in predicting the risk of CRC. Other biases included non-blinding of predictor and outcome assessment, improper use of statistical methods (including the use of univariate analysis for predictor selection and neglecting competing risk), and inadequately small sample sizes. Furthermore, a total of 15 studies raised applicability concerns, which were uniformly linked to the specific selection of participants based on FIT results or their pre-existing cardiovascular disease [41,50,57,61–65,67,72,76].



Figure 3. Risk of bias and applicability assessment of all included studies using the PROBAST (Prediction Model Risk of Bias Assessment Tool) [25].

3.3. Characteristics of Risk Prediction Models for CRC

There are two main categories of risk models for predicting CRC or AN based on their predictor features: (1) conventional risk model, which uses known and well-established risk

factors for CRC; and (2) composite model, which combines well-established risk factors with more advanced features, such as test/laboratory findings and genetic factors. Among the 41 included studies, we identified 22 original conventional risk models, [36–38,40,45,48,50,51,53,54,58,59,61,63–65,67–70,72,74] and 13 composite models [39,41–44,47,56,57,62,71,73,75,76]. Of these, three conventional risk models have undergone external validation in separate studies: (1) the Asia–Pacific Colorectal Screening (APCS) model [45], validated in four other studies [46,49,66,70]; (2) Kaminski’s risk scores [60,69], validated in one study; and (3) The National Cancer Institute—Colorectal Cancer (NCI-CRC) risk assessment tool [77], validated in one study [78]. Eight other studies conducted external validation in the original publication, including six conventional risk models [37,48,63,65,67] and three composite models [41,56,76].

3.3.1. APCS Risk Score

The APCS score is a tool designed to predict the risk of advanced colorectal neoplasia (ACN), which was defined as the presence of colorectal carcinoma or any adenoma at least 10 mm in diameter or with villous histological features or high-grade dysplasia. It uses age, gender, family history of CRC, and smoking status [45]. Following the initial publication, four additional studies expanded on the APCS model, introducing modifications to the data collection design and incorporating extra risk factors [46,49]. Sung et al. (2018) and Luu et al. (2021) externally validated and modified the scores for Hong Kong and Korean populations, adding body mass index (BMI) to the original APCS risk factors, resulting in improved ACN predictions [49,66].

In 2019, He et al. further updated the APCS score for Chinese people over 40 years old, adding BMI, diabetes, and alcohol consumption to the risk factors [46] (Table 3). Later, the Korean Colorectal Screening (KCS) risk model by Hyun Kim et al. was developed by optimizing and adjusting the APCS score for Koreans aged 30 to 74, using age, sex, BMI, family history of CRC, smoking, alcohol, and diabetes as predictors of ACN [70].

3.3.2. Kaminski’s Risk Score

The model was originally derived from a multi-ethnic population aged 40 to 66 in Poland, as part of a national colonoscopy program. The model also aimed to predict the risk of ACN, defined similarly to that in the APCS risk modeling study. However, adenomas were categorized differently in the analysis, with either tubular adenomas or non-neoplastic (Supplementary Materials Supplementary S2). The validated risk index, based on the remainder of the screening cohort ($n = 17,939$ out of 35,918 participants), included age, sex, family history of colorectal cancer, cigarette smoking, and BMI [69]. After the initial publication, Ruco et al. externally validated the Kaminski’s score on Canadians aged 50 to 74 years. This validation revealed that the model was less predictive of ACN in comparison with the original study [60].

3.3.3. The NCI-CRC Risk Assessment Tool

This tool was developed to estimate the 10-year risk of colorectal cancer, while accounting for competing risks such as death [79]. Risk factors for men comprised cancer-negative sigmoidoscopy/colonoscopy in the last 10 years, polyp history in the last 10 years, history of CRC in first-degree relatives, aspirin and NSAID use, cigarette smoking, BMI, current leisure-time vigorous activity (hours per week of intense physical activity), and vegetable consumption. For women, hormone-replacement therapy (HRT) and estrogen exposure based on menopausal status predictors were added. Following the initial publication, the model was externally validated among a veteran cohort undergoing baseline screening colonoscopy [63].

Table 3. Individualized risk prediction models of the included studies.

Study ID (Author and Year)	Target Population Study Period (Cohorts)	All Features Included in the Final Model Name of Risk Prediction Model	AUCs (95% CI)	Calibration P _{HL}
Conventional risk prediction models				
Briggs 2022 [43]	Caucasians, aged 40–80 years 2006–2010 (UK Bank cohorts)	Ethnic group, previous medical history, alcohol use, smoking status, and family history of colorectal cancer, multiple genetic variants, including LDpred2 sparse grid QCancer-10 risk score + LDP-Polygenetic Risk Score (PRS)	Not reported	0.99 ^{a,QM} 0.81 ^{a,QF}
Brand 2017 [36]	Multi-ethnicity, aged < 50 years 2013–2015 (EQUIP-3 study)	Age, sex, BMI, ASA physical status class, ethnicity, and indication (surveillance vs. screening) prediction model for adenoma detection	0.60 (not reported)	1.01 ^a
Cai 2011 [37]	Asians, ≥40 years 2006–2008 (Han citizens)	Age, sex, smoking, diabetes mellitus, green vegetables, pickled food, fried food, and white meat prediction rule for advanced colorectal neoplasm risk	0.74 (0.70–0.78) ¹	0.77 [*]
Cao 2015 [38]	Caucasians, male, aged 40–75 years 1986–2008 (HPFS)	Age, family history of colorectal cancer, BMI, smoking, sitting watching TV/VCR, regular aspirin/NSAID use, physical activity, joint term of multivitamin and alcohol not reported	0.64 (not reported)	0.48 [*]
	Caucasians, female, aged 30–55 years 1986–2008 (NHS)	Age, family history of colorectal cancer, BMI, smoking, alcohol, beef/pork/lamb as main dish, regular aspirin/NSAID, calcium, and oral contraceptive use not reported	0.57 (not reported)	0.96 [*]
Chen 2014 [68]	Asians, aged ≥ 40 years 2011–2012 (AARP-Chinese)	Age, sex, coronary heart disease, egg intake, and defecation frequency risk scoring system for advance colorectal neoplasm	0.75 (0.69–0.82) ¹	0.174 [*]
Deng 2023 [74]	Asians, <50 years 2015–2021 (Fudan cohort, Rejin cohort)	Family history of CRC, smoking, alcohol consumption, processed meat intake, sweet and fried food intake, higher education, eggs and coffee intake, and dietary fiber, calcium, and vitamin supplementation, abdominal discomfort, anorectal symptoms, and intestinal bleeding not reported	0.82 (0.76–0.86) ^b 0.78 (0.74–0.83) ^c	Not reported
He 2019 [46]	Asians, aged > 40 years 2016–2018 (Chinese)	Age, family history of first-degree relatives, smoking alcohol consumption, diabetes, and BMI modified APCS score	0.69 (0.61–0.77)	0.87 ¹
Hong 2017 [51]	Asians, aged ≥ 20 years 2002–2012 (SCS-Korean)	Age, sex, smoking duration, alcohol drinking frequency, and aspirin use not reported	0.71 (0.69–0.74)	Not reported
Hyun Kim 2015 [70]	Asians, aged 30–75 years 2006–2009 (Korean CS)	Age, sex, BMI, family history of colorectal cancer, smoking, alcohol, and diabetes KCS score	0.68 (0.61–0.76) ¹	0.48 [*]

Table 3. Cont.

Study ID (Author and Year)	Target Population Study Period (Cohorts)	All Features Included in the Final Model Name of Risk Prediction Model	AUCs (95% CI)	Calibration P _{HL}
Conventional risk prediction models				
Imperiale 2015 [53]	Caucasians, aged 50–80 years 2004–2011 (not reported) §	Age, sex, family history of CRC, cigarette smoking, and waist circumference not reported	0.72 (not reported) ¹	0.42 *
Imperiale 2021 [52]	Caucasians, aged 50–80 years 2004–2011 (not reported) §	Age, sex, marital status, education, smoking, significant ethanol use, NSAID use, aspirin use, metabolic syndrome, red meat consumption, regular activity (10 years), moderate activity (over last years), and vigorous activity (over last year) not reported	0.78 (not reported) ¹	0.37 * (0.69 * in validation set)
Jung 2017 [72]	Asians, aged 30–49 years 2010–2014 (Korean- SHS)	Age, sex, BMI, family history of colorectal cancer, and smoking habits Probability of Advanced colorectal neoplasia in a population aged < 50 years (PAC-50)	0.67 (0.65–0.70)	0.093 *
Jung 2018 [61]	Asians (FIT-negative), aged ≥ 40 years 2010–2014 (Korean-SHS)	Age, current smoker, overweight, obesity, hypertension, and old cerebrovascular attack risk scores in fit-negative participants	Not reported	Not reported
Kaminski 2014 [69]	Multi-ethnicity, aged 40–66 years 2007 (Poland NCSP)	Age, sex, family history of colorectal cancer, and cigarette smoking, BMI Kaminski’s risk prediction model	0.62 (0.60–0.64) ¹	1.0 ^a
Kim 2019 [50]	Asians, aged < 50 years 2003–2012 (Koreans)	Age, sex, alcohol, smoking, BMI, glucose metabolism abnormality, and family history of colorectal cancer YCS score	0.66 (0.65–0.67) ¹	0.261 *
Liu 2018 [67]	Caucasians, Male, aged 40–75 years 1986–2010 (HPFS US-based)	Age, higher BMI, more pack-years of smoking, higher alcohol consumption with lower levels of multivitamin use, family history of colon cancer, and colonoscopy or sigmoidoscopy screening not reported	0.62 (not reported)	1.05 ^a
	Caucasians, female, aged 30–55 years 1986–2010 (NHS US-based)	Age, height, BMI, postmenopausal hormone use, physical activity, pack-years of smoking, calcium intake, alcohol and multivitamin intake, aspirin use, history of CRC, family history, and colonoscopy not reported	0.60 (not reported)	1.19 ^a
Luu 2021 [66]	Asians, aged ≥ 40 years 2002–2014 (Korean Cancer Screening)	Age, sex, first-degree family history of CRC, and smoking status APCS score	0.62 (not reported) ^c	Not reported
Ma 2010 [65]	Asians, male, aged 40–69 years 1993–2005 (JPHC and PHC, Japanese)	Age, BMI, daily physical activity, alcohol consumption, smoking habit, family history of colorectal cancer, and diabetes diagnosis JPHC risk prediction model	0.70 (0.68–0.72)	0.08 *

Table 3. Cont.

Study ID (Author and Year)	Target Population Study Period (Cohorts)	All Features Included in the Final Model Name of Risk Prediction Model	AUCs (95% CI)	Calibration P _{HL}
Conventional risk prediction models				
Murchie 2017 [64]	Multi-ethnicity, aged 40–59 years 2008–2014 (not reported)	Age, sex, BMI, and smoking history calculator for high-risk colon adenomas	0.64 (not reported)	Not reported
Musselwhite 2019 [63]	Multi-ethnicity, aged 50–75 years 1994–1997 (Veterans)	Age, history of colonoscopy or endoscopy in the last 10 years, whether polyps were observed, family history of CRC, weekly physical activity, aspirin or NSAIDs use, smoking, vegetable intake, and BMI NCI-CRC risk assessment tool (external validation)	0.60 (0.57–0.63) ³	Not reported
Ruco 2015 [60]	Caucasians, aged 50–74 years 2003–2008 (VACS, US-based)	Age, sex, family history of CRC, smoking history, and BMI Kaminski’s risk prediction model	0.64 (0.61–0.67)	Not reported
Schroy III 2015 [54]	Multi-ethnicity, aged 50–79 years 2009 (BMC)	Age, sex, smoking, alcohol intake, height, and combined sex/race/ethnicity risk prediction index for advanced colorectal neoplasia	0.69 (0.66–0.72) ¹	0.73–0.93 *
Sekiguchi 2018 [59]	Asians, aged ≥ 40 years 2004–2013 (NCC, Japanese)	Sex, age, first-degree relatives with CRC, BMI, and smoking history de novo risk score for advanced colorectal neoplasia	0.70 (0.67–0.73)	0.71 *
Sharara 2020 [58]	Lebanese, aged ≥ 50 years Not reported (AUBMC patients)	Age, BMI, smoking status, and daily consumption of red meat not reported	0.73 (0.66–0.79)	0.85 *,a
Shin 2014 [48]	Asians, male, aged 30–80 years 1996–1999 (NHIC, Koreans)	Age, BMI, serum cholesterol, family history of cancer, and alcohol consumption not reported	0.76 (0.75–0.77) ^b	1.29 ^a
	Asian, female, aged 30–80 years 1996–1999 (NHIC, Koreans)	Age, height, and meat intake frequency not reported	0.71 (0.70–0.72) ^b	1.23 ^a
Sung 2018 [49]	Asians, aged 50–70 years 2008–2012 (Hong Kong-based)	Age, sex, BMI, family history of CRC, and smoking history APCS score	0.65 (0.61–0.69)	0.57 *
Sutherland 2021 [40]	Caucasians, aged 50–74 years 2008–2016 (CCSC)	Age, sex, BMI, smoking status, diabetic status, family history of CRC, alcohol consumption, and vitamin D supplementation not reported	0.69 (0.65–0.72) ⁴	Not reported
Tao 2014 [55]	Germans, ≥50 years 2005–2011 (BliTz study)	Age, sex, first-degree relatives with a history of CRC, cigarette smoking, alcohol consumption, red meat consumption, ever regular use of NSAIDs, previous colonoscopy, and previous detection of polyps de novo risk prediction model	0.67 (0.65–0.69) ^{b,1} 0.71 (0.67–0.75) ^{b,2}	Not reported

Table 3. Cont.

Study ID (Author and Year)	Target Population Study Period (Cohorts)	All Features Included in the Final Model Name of Risk Prediction Model	AUCs (95% CI)	Calibration P _{HL}
Conventional risk prediction models				
Yeoh 2011 [45]	Asians, mean age of 54.4 years (SD ± 11.6 years), 2004 (Multi-ethnic cohorts from Asia)	Age, sex, first-degree family history of CRC, and smoking status APCS score	0.64 (0.57–0.71) ¹	0.49 ^{a,b}
Composite risk prediction models				
Arnau-Collell 2022 [75]	Hispanic, aged 50 to 69 years 2009 = 2019 (CRC screening cohorts)	Sex, age, FIT value, and polygenic risk score	0.64 (0.61–0.66)	Not reported
Auge 2014 [73]	Spanish/Catalan, aged 50–69 years 2009–2012, (Barcelona CRC screening Program, FIT-positive individuals)	Age, sex, and FHbC result not reported	0.67 (not reported)	0.31 [*]
Cooper 2020 [71]	UK patients, aged 60–74 years 2009–2017 (BCSS)	Age, sex, smoking status, alcohol consumption (units per week), previous negative FOBT, family history of gastrointestinal cancer, and FOBT result not reported	0.86 (0.85–0.87)	Not reported
Meester 2022 [42]	Dutch, aged 55 to 75 years 2014–2019 (Dutch CRC Screening cohort)	Age, sex, first and second FHbC not reported	0.78 (0.77–0.79) ¹ 0.73 (0.71–0.75) ²	0.98–0.99 ^a
Müdler 2023 [44]	Dutch, aged 55 to 75 years 2014–2021 (Dutch citizens)	Age, sex, f-Hb previous round, and the two most recent f-Hb concentrations not reported	0.79 (0.78–0.80) ^{b,1} 0.76 (0.74–0.78) ^{b,2}	Not reported
Park 2019 [62]	Asians, aged ≥ 50 years 2013–2017 (NCSP, Koreans)	Age, sex, smoking habit, obesity, diabetes mellitus, and FHbC not reported	0.90 (0.86–0.93)	0.26 [*]
Soonklang 2021 [57]	Asians, aged 50–65 years 2009–2010 (Thai)	Age, sex, BMI, family history of CRC in first-degree relatives, smoking, diabetes mellitus, and FIT result not reported	0.77 (0.71–0.84)	Not reported
Stegeman 2014 [39]	Dutch, aged 50–75 years 2009–2010 (COCOS)	Age, sex, total calcium intake, family history of CRC, number of family member with CRC, alcohol intake, smoking history, regular use of aspirin or non-steroid anti-inflammatory drug (NSAID), physical activity, and FIT result not reported	0.76 (not reported)	0.94 [*]
Thomsen 2022 [41]	Danish residents, aged 50–74 years 2014–2016 (DCCSD and DCCG)	Age, sex, and FIT result not reported	0.67 (0.67–0.68) ^{c,1} 0.75 (0.74–0.76) ^{c,2}	1.02 ^{a,1} 0.99 ^{a,2}
Van 't Klooster 2020 [76]	Dutch, aged 45–80 years 2005–2012 (UCC-SMART and CANTOS)	Age, sex, smoking, weight, height, alcohol use, antiplatelet use, diabetes, and C-reactive protein not reported	0.64 (0.58–0.70) ^{c,2}	0.85 ^{c,d}

Table 3. Cont.

Study ID (Author and Year)	Target Population Study Period (Cohorts)	All Features Included in the Final Model Name of Risk Prediction Model	AUCs (95% CI)	Calibration P _{HL}
Composite risk prediction models				
Yang 2017 [47]	Asians, aged 50–70 years 2003–2012 (Koreans)	Age, sex, family history of colorectal cancer, smoking, BMI, serum levels of fasting glucose, low-density lipoprotein cholesterol, and carcinoembryonic antigen Samsung Colorectal Screening (SCS) risk model ***	0.68 (0.67–0.69)	0.35 *
Yen 2014 [56]	Asians, aged ≥ 40 years 2001–2007 (KCIS, Taiwanese)	Sex, FHbC result, family history of colorectal cancer, type 2 diabetes, hypertension, alcohol drinking, smoking, BMI, triglyceride level, total cholesterol not reported	0.63 (0.62–0.65) ^{c,+} 0.86 (0.85–0.87) ^{c,++}	Not reported

Cohort abbreviations: HPFS, Health Professional Follow-Up Study; NHS, Nurses' Health Study; AARP, asymptomatic average-risk population; KHSH, Kangbuk Samsung Health Study; JPHC, Japan Public Health Center-based prospective study; PHC, public health center-based areas; VACS, Veterans Affairs Cooperative Study; NCSP, National Cancer Screening Programme; COCOS Study, Colonoscopy or Colonography for Screening Study; NHIC, National Health Insurance Corporation; BCSS, Bowel Cancer Screening Programme; UCC-SMART, Utrecht Cardiovascular Cohort—Second Manifestations; DCCSD, Danish Colorectal Cancer Screening Database; DCCG, Danish Colorectal Cancer Group Database; CANTOS, Canakinumab Anti-Inflammatory Thrombosis Outcomes Study; ACNR, Advanced colorectal neoplasia risk; KCIS, Keelung community-based integrated screening; BMC, Boston Medical Center; NCC, National Cancer Center in Tokyo, Japan; CCSC, Colon Cancer Screening Centre; NCSP, National Colonoscopy Screening Program. Features and model abbreviations: FIT, fecal immunochemical test; FOBT, Faecal Occult Blood Test; FHbC, fecal hemoglobin concentration; BMI, body mass index; + conventional risk model only; ++ conventional risk in combination with a test/laboratory findings; APCS, Asia–Pacific Colorectal Screening risk model; KCS, Korean Colorectal Screening risk model; YCS, Young Colorectal Screening; *** SCS c-statistics was found superior to the APCS and KCS risk models; NCI, National Cancer Institute; ^{QM}, QCancer-10 model for men; ^{QF}, QCancer-10 model for female. Outcome measures: CRC, colorectal cancer; AN, advanced neoplasia; ¹ advanced neoplasms diagnosis at colonoscopy; ² CRC diagnosis at colonoscopy; ³ risk of AN at 5 years; ⁴ high-risk adenomas; ⁵ cumulative risk of colon cancer up to 70 years. Statistical terms: CI, confidence interval; P_{HL}, Hosmer–Lemeshow goodness-of-fit test; * *p*-value higher than 0.05 indicating adequate calibration; ^a observed/expected ratio or calibration slope; ^b development or derivation dataset; ^c external validation dataset; ^d the expected-to-observed ratios of the event of interest and competing event; §, model updating.

3.4. Clustering Analysis among Risk Prediction Models

The hierarchical clustering analysis of 44 risk models generated five distinct clusters and 16 nodes based on five variable domains, namely (1) sociodemographic factors (e.g., age, gender, and race), (2) behavioral/dietary factors (e.g., smoking habits, alcohol use, physical activities, and meat intake), (3) clinical factors (e.g., NSAID/aspirin use, BMI, diabetes, and hypertension), (4) patient/family history (e.g., family history of history and indication screening or surveillance), (5) test/laboratory results (e.g., FhBC and FIT), and (6) biomarkers (e.g., carcinoembryonic antigen and C-reactive protein) (Figure 4).

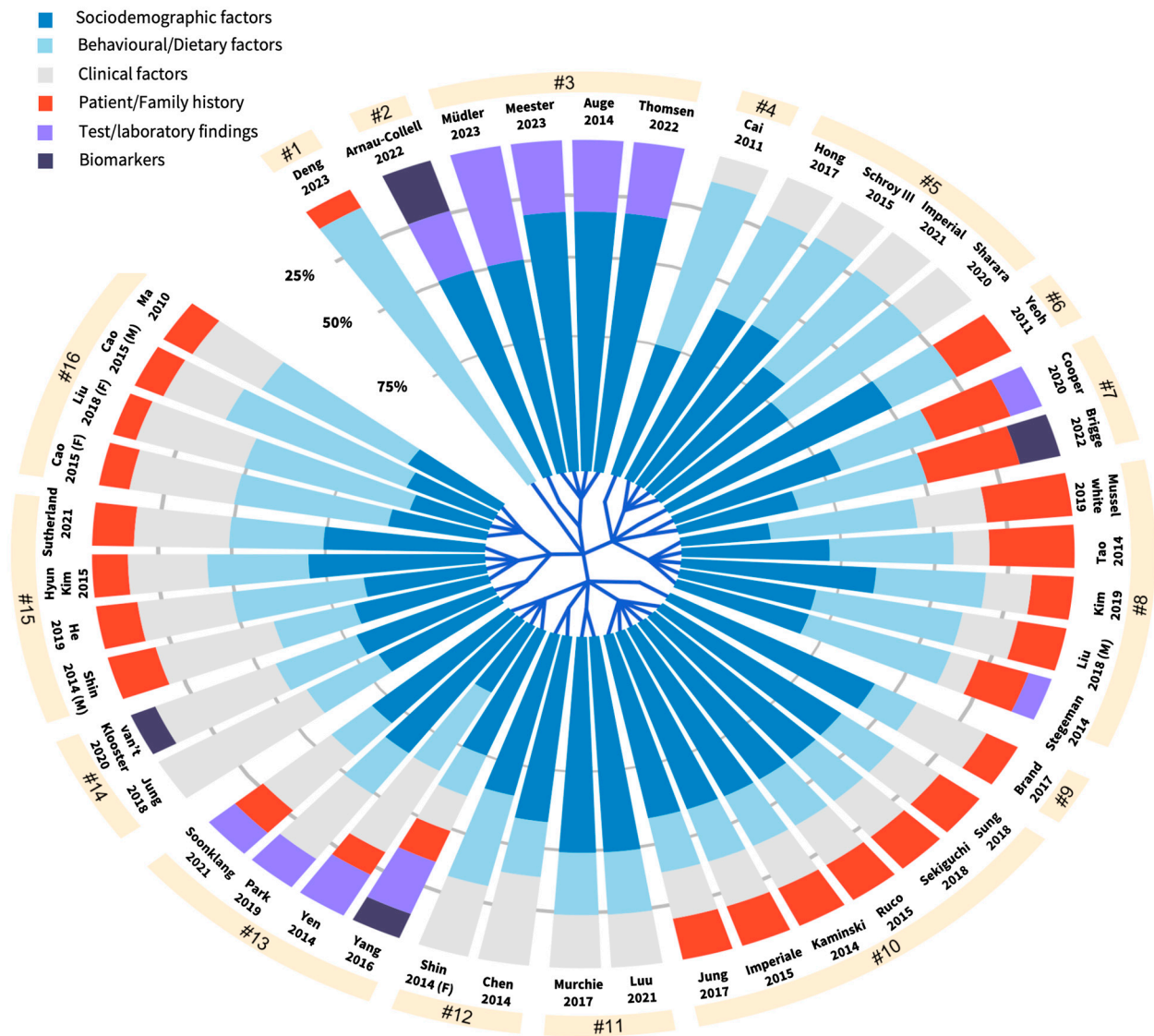


Figure 4. Hierarchical clustering of 44 risk prediction models (n = 41 studies) based on five variable domains.

Each slice of the chart represents one risk prediction model. The sectors in each study indicate what types of domains were included in each model, along with their corresponding proportions. The blue branches starting from the center of the chart show how the tools were divided based on the similarity of their domain profiles. For instance, the second cluster (sub-node #3), which consists of four models, exhibited strikingly similar characteristics. The majority of the predictors used sociodemographic factors and were combined with test/laboratory findings. Specifically, the FIT test results served as predictors of CRC or AN. Meanwhile, the third and fourth clusters demonstrate high

heterogeneity. Specifically, within the 7th and 13th node, the risk prediction models integrate at least four domains, namely sociodemographic factors, behavioral factors, clinical factors and test/laboratory findings, and biomarkers.

3.5. Risk Predictors for Asian Populations

The newly developed risk prediction models between 2011 and 2021 predominantly originated from Asian countries ($n = 19$), including China [37,46,68,74], Korea [47,48,50,51,61,62,66,70,72], Japan [59,65], Thailand [57], Hong Kong [45,49], and Taiwan [56]. Among these novel risk prediction models based on Asian populations, the most common predictors were age ($n = 19$), smoking history ($n = 17$), sex ($n = 15$), BMI ($n = 14$), family history of CRC ($n = 14$), alcohol intake ($n = 8$), and diabetes mellitus ($n = 6$) [37,45–51,56,57,59,61,62,65,66,68,70,72,74,80].

3.6. Risk Predictors for Caucasian Populations

Several studies reported different models, published between 2014 and 2022, based on Caucasian cohorts from the United States [52,53,67,76], Canada [40], the United Kingdom [71], Germany [55], Denmark [41], and the Netherlands [39]. Among these models, the most common predictors were age ($n = 11$), sex ($n = 10$), smoking history ($n = 8$), alcohol intake ($n = 6$), and family history of CRC ($n = 6$) [39–42,44,52,53,55,67,71,76]. On the other hand, predictors that were less common included BMI ($n = 2$) [40,67], NSAID use ($n = 3$), physical activity ($n = 3$) [39,52,55], red meat consumption ($n = 2$) [52,55], and height ($n = 1$) [76].

3.7. Risk Predictors for Multi-Ethnic or Ethnic Minority Populations

Eight studies reported models that were derived from multi-ethnic groups, including Black, Hispanic or Latino, Asian, Afro Caribbean, Lebanese, and Afro American from Poland [69], Spain [73,75], the United States [36,54,63,64], and Lebanon [58]. Among these models, the most common predictors were age ($n = 7$), sex ($n = 6$), smoking history ($n = 5$), and BMI ($n = 5$) [36,54,58,63,64,69,75]. Additionally, ethnicity ($n = 2$) [36,54], height ($n = 1$), and alcohol intake ($n = 1$) [54] were also identified as significant predictors of CRC or AN. To note, weekly physical activity, the use of aspirin or NSAIDs, and vegetable consumption emerged as distinctive protective predictors that were exclusively reported in a model derived from the cohorts of veterans in the United States [63].

3.8. Performance of Risk Prediction Models for CRC

All included studies, except for that of Jung et al. (2018) and Briggs et al. (2023), reported AUC values to assess the discriminatory accuracy of the models (Table 3). Among the 44 included models, 23 conventional risk models [37,40,42,44–46,48–51,54,58–61,63–65,67–70,72,74,75] and 10 composite models [39,41,47,56,57,62,71,75,76,80] reported AUC values with a 95% confidence interval (CI), as presented in Figure 5. Meanwhile, the model calibration was assessed in two different metrics: using the expected-to-observed (E/O) ratios, which were only demonstrated in nine of the studies; or by employing the Hosmer–Lemeshow (HL) goodness-of-fit test, as was the case in 12 out of 41 studies.

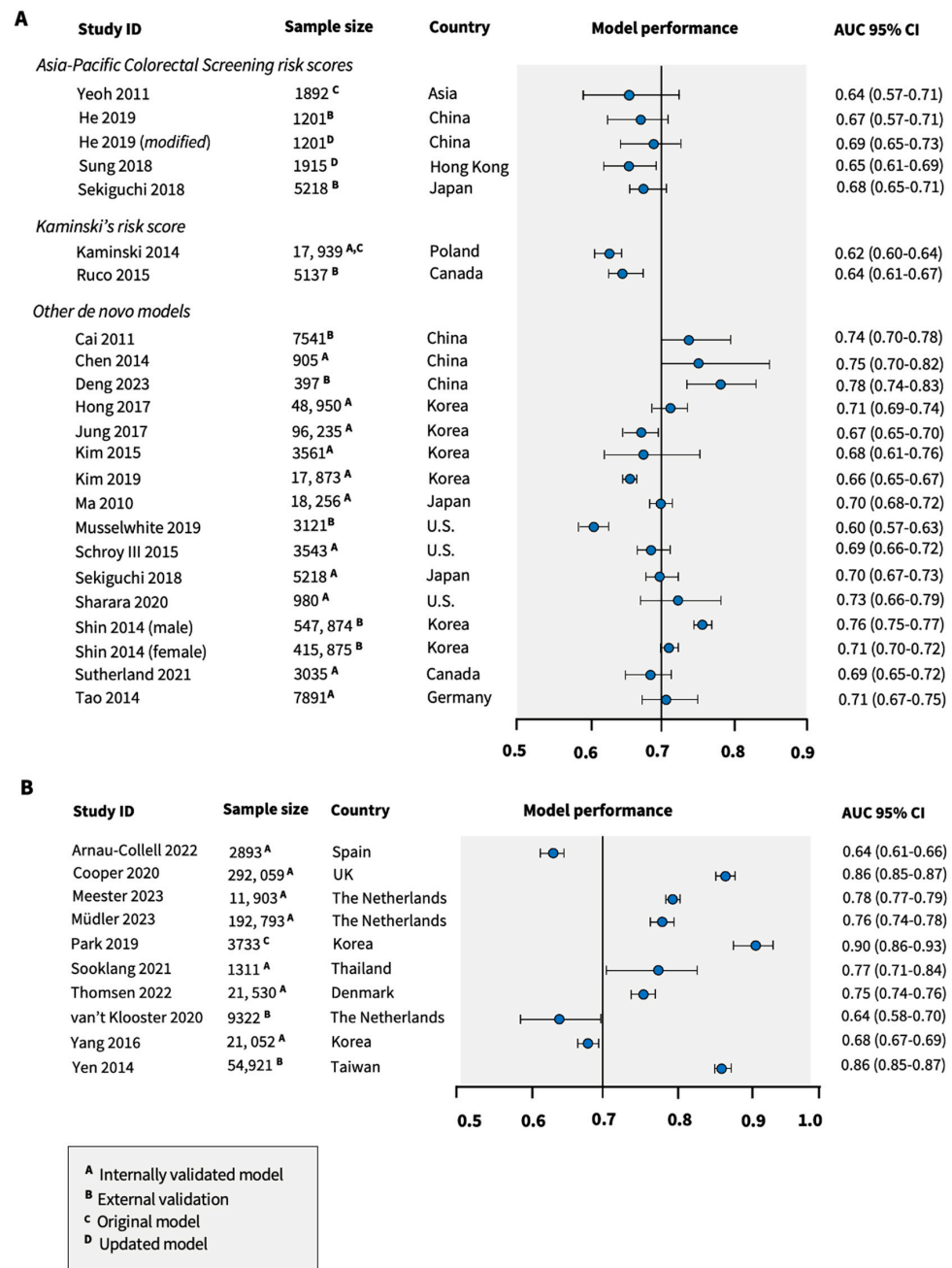


Figure 5. Model performance of (A) conventional [37,40,42,44–46,48,50,51,54,55,58–60,63,65,68–70,72,74,75,78] and (B) composite risk models for colorectal cancer with 95% confidence interval (CI) [41,42,44,47,56,57,62,71,75,76].

3.8.1. Performance of the (Modified) APCS Risk Score

The APCS score, through various modifications, exhibited an enhanced performance [45,46,49,66,70]. Initially, the score had an AUC of 0.64 (95% CI: 0.57–0.71) with good calibration (HL goodness-of-fit p -value (P_{HL}) = 0.49) [45]. However, when externally validated with a Korean cohort, it demonstrated a lower AUC of 0.62 compared with the original study and without calibration details [66]. In contrast, the modified APCS score of Sung et al. showed a slight AUC improvement to 0.65 (95% CI: 0.61–0.69), with good calibration (P_{HL} = 0.57) [49] (Figure 5A). In another external validation study by He et al., the addition of BMI, diabetes, and alcohol consumption led to an AUC of 0.69 (95% CI: 0.65–0.73) and a good calibration (P_{HL} = 0.87) [46]. This modified version, combined with

FIT results, achieved an NPV of 98%, marking an advancement over just FIT (NPV of 97%) or the APCS score alone (NPV of 97.9%) [46].

3.8.2. Performance of the Kaminski's Risk Score

Kaminski's risk scores, originally based on multi-ethnic groups from Poland, had an AUC of 0.62 (95% CI: 0.60–0.64) [60,69]. This model showed good calibration, with a test dataset P_{HL} of 0.74 and a validation dataset P_{HL} of 0.16, along with an E/O ratio of 1.00 (95% CI: 0.95–1.06) [69]. Subsequently, when externally validated on a large Canadian cohort, the AUC slightly increased to 0.64 (95% CI: 0.61–0.67). The PPV of the model was found to be 5.88%, indicating a limited ability to accurately identify high-risk individuals for CRC. In contrast, its NPV was 93.16%, thus effectively classifying those without CRC as low risk [60] (Supplementary Materials Supplementary S4).

3.8.3. Performance of Models Based on Asian Populations

Deng et al. (2023), who focused on a Chinese population, reported the highest AUC of 0.78 (95% CI: 0.74–0.83) [48] (Figure 5A). This high value was attributed to the inclusion of lifestyle and dietary factors, clinical symptoms, and medical and family history in their prediction model, which could be closely linked to the occurrence of CRC. Other models from Chen et al. and Cai et al., also based on Chinese cohorts, achieved AUCs of 0.75 (95% CI: 0.70–0.82) and 0.74 (95% CI: 0.70–0.78), respectively [37,68]. Models from Hong et al., Ma et al., and Shin et al. exhibited AUCs ranging between 0.70 and 0.71 [48,51,65]. Sekiguchi et al.'s model, which is a Japanese cohort-based model, displayed a better performance (AUC = 0.71 (95% CI: 0.67–0.73) compared with the APCS score (AUC = 0.68; 95% CI: 0.65–0.71) [59]. On the other hand, the KCS score only exhibited modest discriminatory accuracy, with an AUC of 0.68 (95% CI: 0.61–0.75) [70], but suggested good calibration from both datasets: the derivation dataset (P_{HL} = 0.81) and the validation dataset (P_{HL} = 0.48) (Table 3).

For composite models, the model developed by Park et al. had the highest reported AUC of 0.90 (95% CI: 0.86–0.93), indicating an almost excellent discriminatory accuracy [62] (Figure 5B). Similarly, the model developed by Yen et al. (2014), based on Taiwanese populations, also demonstrated a high discriminatory accuracy of 0.86 (95% CI: 0.85–0.87). It is worth noting that both models incorporated FhbC results, and the one by Yen et al. also included a triglyceride-level predictor.

3.8.4. Performance of Models Based on Caucasian Populations

Of 15 models based on Caucasian populations, the model developed by Imperiale et al. in 2021 had the highest AUC (0.78) and exhibited good calibration for both the derivation (P_{HL} = 0.37) and validation (P_{HL} = 0.69) datasets [52]. Other notable models include Stegeman et al. (AUC = 0.76) [39], Imperiale et al. (2015) (AUC = 0.72, 95% CI: not reported) [53], and Tao et al. (AUC = 0.71, 95% CI: 0.67–0.75) [55] (Figure 5A). Only Imperiale et al. (2015) reported the model calibration (P_{HL} = 0.42). The model by Cao et al. had the lowest AUC of 0.57 (95% CI was not reported) but exhibited good calibration (P_{HL} of 0.48) (Table 3). Interestingly, the NCI-CRC tool evaluated risks of CRC at 5, 10, and 20 years for the veteran cohort, showing a declining AUC over time (0.60 to 0.58) [63]. However, the study did not report model calibration, which limits the validity of the prediction model.

For the composite risk models based on Caucasians, the model by Cooper et al. achieved the highest AUC (0.86, 95% CI: 0.85–0.87) by integrating the FOBT results, along with conventional risk predictors, but lacked model calibration. In contrast, only the model derived by van 't Klooster exhibited modest discriminatory accuracy, with an AUC of 0.64 (95% CI: 0.58–0.70), and it displayed good calibration (P_{HL} = 0.85) [76] (Table 3 and Figure 5B).

3.8.5. Performance of Models Based on Multi-Ethnic or Ethnic Minority Populations

Five out of seven models derived from multi-ethnic groups used conventional risk predictors to estimate the risk of CRC or ACN [36,54,58,63,64]. Sharara et al.'s model showed the highest accuracy (AUC = 0.73) [58], while the other five models had moderate

discriminatory accuracy (AUCs between 0.60 and 0.69) [36,54,58,63,64]. Additionally, four conventional risk models demonstrated good calibration, either using the P_{HL} ($n = 2$) [54,58] or the E/O ratios ($n = 2$) [36,69].

3.9. Generalizability, Clinical Utility, and Usability of the Models

3.9.1. Generalizability

When examining generalizability, we observed that most of the included studies lacked model validation in an independent cohort ($n = 27$), while others excluded relevant individuals eligible for CRC screening ($n = 14$) (Table 4). Moreover, among the development studies that only conducted internal validation, only nine performed either bootstrapping or 10-fold cross-validation, while 12 employed a random split-sampling approach, and four had unclear information regarding the internal validation approach.

Table 4. Summary of evidence of the model's performance, validity, and potential clinical utility.

Reported Metrics	Out of 41 Studies n (%)	References
AUCs reported	39 (95.1)	[36–42,44–60,62–76]
Expected/observed ratios	9 (21.9)	[36,38,41,48,65,67,69,71,76]
Sensitivity	16 (39.0)	[37,39,40,43,46,50,51,60,61,66,68,70–72,74,75]
Specificity	15 (36.6)	[37,39,40,43,50,51,60,61,66,68,70–72,74,75]
Net reclassification improvement	3 (7.3)	[39,54,67]
Negative predictive value	6 (14.6)	[40,46,51,60,68,75]
Positive predictive value	6 (14.6)	[40,51,60,68,71,75]
Number needed to screen for CRC or number needed to refer for colonoscopy	5 (12.2)	[37,49,55,68,72]
Appropriate handling of continuous variables	18 (43.9)	[36,38,40,42–44,48,51,53,58,62–64,67,69,71,75,76]
Risk threshold determination	27 (65.8)	[37,39–42,44–47,49–55,59,60,62–66,69–73,75]
■ Risk threshold arbitrarily	14/27 (51.8)	[37,42,45–47,50,53,55,61,62,66,70,73,75]
External validation studies	13 (31.7)	[37,41,46,48,49,56,60,63,65–67,74,76]
Clinical usable models	26 (63.4)	[38,40–42,44–47,49,53–55,57,60,62–68,70,71,73,75,76]
■ Simple risk score/risk assessment tool	8 (19.5)	[38,45,46,49,63,65–67]
Potential clinical utility		
⊕⊕⊕⊕	5 (12.2)	[39,40,60,71,76]
⊕⊕⊕	9 (21.9)	[37,41,43,44,48,51,54,66,75]
⊕⊕	18 (43.9)	[38,42,46,47,49,52,53,55,59,63,65,67–70,72–74]
⊕	9 (21.9)	[36,45,50,56–58,61,62,64]

⊕⊕⊕⊕, High potential: interpretable performance metrics, risk estimates, decision curves/net benefits, and (multiple) plausible risk thresholds were reported. Appropriate handling of continuous predictors and risk threshold determinations were observed. ⊕⊕⊕, Some concerns: minor issues exist in terms of performance metrics reporting or risk threshold determination, including few (≤ 2) missing performance metrics and use of arbitrary risk thresholds, respectively. ⊕⊕, Low potential: limited potential for clinical utility due to inappropriate handling of continuous predictors, unclear model calibration, risk threshold determination, sensitivity, specificity, and other estimates demonstrating clinical utility. ⊕, Very low potential: very minimal information provided, making it hard to assess any meaningful clinical utility.

We further identified the most common pitfalls related to model development and validation that could potentially influence the generalizability. Notably, several studies failed to report calibration using E/O ratios and goodness-of-fit (PHL) measurements and did not include calibration plots to check for alignment of predictions in actual CRC or AN outcome. This lack of reporting hinders our comprehension of the model stability and the distribution of risk estimates among all individuals in a dataset. Other concerns included the use of split-sample techniques for internal validation, reliance on univariate analysis for risk predictor selection, a lack of reporting on the net reclassification index, and improper handling of continuous variables like age and BMI without a sufficient basis (Table 5).

Table 5. Evidence matrix for evaluating the potential clinical utility of the model.

Study ID	Sample Size (n)		Prevalence of ACN (%)	Expected/Observed Ratio Reported	Model's Sensitivity Reported	Model's Specificity Reported	Appropriate Handling of Continuous Predictors †	Risk Threshold Determination	Other Estimates Demonstrating Clinical Utility Reported	Potential Clinical Utility
	DC (n)	VC (n)								
Asia-Pacific Cancer Screening (APCS) risk prediction model ***										
Yeoh 2011 [45] ++	860	1892	4.5 ¹ 3.0 ²	No	No	No	No	Arbitrarily determined	No	⊕
Sung 2018 [49] + + +	3829	1915	5.4 ¹ 6.0 ²	No	No	No	No	Prevalence as threshold	Yes	⊕⊕
He 2019 [46] + + +	995	1201	4.1 ¹ 3.7 ²	No	Yes	No	No	Arbitrarily determined	Yes	⊕⊕
Luu 2021 [66] + + +	12,520	-	2.5	No	Yes	Yes	No	Arbitrarily determined	Yes	⊕⊕⊕
Kaminski's risk prediction model										
Kaminski 2013 [69] + +	17,979	17,939	7.1	Yes	No	No	Yes	Prevalence as threshold	No	⊕⊕
Ruco 2015 [60] + + +	-	5137	6.8	No	Yes Range	Yes Range	No	Prevalence as threshold	Yes	⊕⊕⊕⊕
Other risk prediction model with external validation										
Cai 2012 [37] + + +	5229	2312	6.4	No	Yes	Yes	No	Arbitrarily determined	Yes	⊕⊕⊕
Deng 2023 [74] + + +	1087	397	NA	No	Yes	Yes	No	Unclear	Yes	⊕⊕
Liu 2018 [67] + + +	103,249	-	1.12	Yes	No	No	Yes	Unclear	Yes	⊕⊕
Ma 2010 [65] + + +	28,115	18,256	1.9 ¹ 2.2 ²	Yes	No	No	No	Based on absolute risk	No	⊕⊕
Musselwhite 2019 [63] + + +	3121	-	11.0	No	No	No	Yes	Based on absolute risk	No	⊕⊕
Shin 2014 [48] + + +	1,326,058	963,749	0.69	Yes	No	No	Yes	Not reported	No	⊕⊕⊕
Thomsen 2022 [41] + + +	34,929	21,530	5.9	Yes	No	No	No	Prevalence of FIT positive as threshold	No	⊕⊕⊕
van 't Klooster 2020 [76] + + +	7280	9322	2.5	Yes	No	No	Yes	NA	Yes	⊕⊕⊕⊕
Yen 2014 + + +	54,921	Unclear	-	No	No	No	No	Not reported	No	⊕
De novo models without external validation										
Arnau-Collell 2022 [75] + +	2893	-	NA	No	Yes	Yes	Yes	Arbitrarily determined	No	⊕⊕
Auge [73] + +	3109	-	9.5	No	No	No	No	Arbitrarily determined	Yes	⊕⊕
Brand 2017 [36] + +	9934	10,034	40 ●	Yes	No	No	Yes	Not reported	No	⊕
Briggs 2022 [43] + +	30,000	280,664	1.5	No	Yes	Yes	Yes	Utility-based risk threshold	Yes	⊕⊕⊕⊕
Cao 2015 [38] + +	17,970 W 4881 M	§	3.8 W 6.7 M	Yes	No	No	Yes	Unclear	No	⊕⊕
Chen 2014 [68] + +	905	§	5.3	No	Yes	Yes	No	Unclear	Yes	⊕⊕
Cooper 2020 [71] + +	292,059	§	5.41	Yes	Yes	Yes	Yes	Threshold minimizing misclassification	Yes	⊕⊕⊕⊕
Hong 2017 [51] + +	24,725	24,725	2.3	No	Yes	Yes	Yes	Prevalence as threshold	Yes	⊕⊕⊕
Hyun Kim 2015 [70] + +	2152	1316	4.4	No	Yes	Yes	No	Arbitrarily determined	No	⊕⊕
Imperiale 2015 [53] + +	2993	1467	9.4	No	No	No	Yes	Arbitrarily determined	No	⊕⊕
Imperiale 2021 [52] + +	3025	1475	9.1	No	No	No	No	Prevalence as threshold	No	⊕⊕
Jung 2017 [72] + +	57,635	38,600	1.3	No	Yes	Yes	No	Based on Youden index	Yes	⊕⊕
Jung 2018 [61] + +	11,873 FIT-	-	2.1	No	Yes	Yes	No	Arbitrarily determined	No	⊕
Kim 2019 [50] + +	41,702	17,873	0.9	No	Yes	Yes	No	Arbitrarily determined	No	⊕
Meester 2022 [42] + +	266,881	11,903	1.2 AN 0.2 CRC	Yes	No	No	Yes	Prevalence as threshold	Yes	⊕⊕⊕
Müdler 2023 [44] + +	219,258	192,793	1.7	No	No	No	Yes	Prevalence as threshold	Yes	⊕⊕⊕
Murchie 2017 [64] + +	5063	§	5.7	No	No	No	Yes	Unclear	No	⊕
Park 2019 [62] + +	3733	-	9.8	No	No	No	Yes	Arbitrarily determined	No	⊕
Schroy III 2015 [54] + +	3543	§	5.7	No	No	No	No	Based on predicted probability	Yes	⊕⊕⊕
Sekiguchi 2018 [59] + +	5218	§	4.3	No	No	No	No	Threshold minimizing misclassification	No	⊕⊕⊕
Sharara 2020 [58] + +	980	§	5.10	No	No	No	Yes	Not reported	No	⊕
Soonklang 2021 [57] + +	1311	§	4.04	No	No	No	No	Not reported	No	⊕
Stegeman 2014 [39] + +	1121	-	9.1	No	Yes	Yes	No	Utility-based risk threshold determination	Yes	⊕⊕⊕⊕
Sutherland 2021 [40] + +	3035	§	7.53	No	Yes	Yes	Yes	Based on predicted probabilities.	Yes	⊕⊕⊕⊕

Table 5. Cont.

Study ID	Sample Size (n)		Prevalence of ACN (%)	Expected/Observed Ratio Reported	Model's Sensitivity Reported	Model's Specificity Reported	Appropriate Handling of Continuous Predictors †	Risk Threshold Determination	Other Estimates Demonstrating Clinical Utility Reported	Potential Clinical Utility
	DC (n)	VC (n)								
Tao 2014 [55] ++	7891	3519	9.9	No	No	No	No	Arbitrarily determined	Yes	⊕⊕
Yang 2016 [47] ++	49,130	21,052	1.4	No	No	No	No	Arbitrarily determined	No	⊕⊕

+ Development study; ++ development study with internal validation; +++ (with) external validation study; † in the context of individualized risk prediction model; ● refers to the prevalence of any colorectal adenoma that is histologically confirmed; ¹, derivation cohort; ², internal validation cohort; ^M, Men; ^W, Women; ^{FIT}, fecal immunochemical test; ^{AN}, Advanced Neoplasia; ^{CRC}, Colorectal cancer; § is used 10-fold cross-validation or bootstrapping method; ^{Range}, present range of scenarios across scores (with different sensitivity and specificity/multiple plausible threshold); *** existing effectiveness studies of the APCS risk scoring-system combined with FIT results have been reported by Aniwan et al. (2015) and Chiu et al. (2016). ⊕⊕⊕⊕, High potential: interpretable performance metrics, risk estimates, and (multiple) plausible risk thresholds were reported. Appropriate handling of continuous predictors and risk threshold determinations were observed. ⊕⊕⊕, Some concerns: minor issues exist in terms of performance metrics reporting or risk threshold determination, including few (≤ 2) missing performance metrics, and use of arbitrary risk thresholds, respectively. ⊕⊕, Low potential: limited potential for clinical utility due to inappropriate handling of continuous predictors, unclear model calibration, risk threshold determination, sensitivity, specificity, and other estimates demonstrating clinical utility. ⊕, Very low potential: very minimal information provided, making it hard to assess any meaningful clinical utility.

3.9.2. Potential Clinical Utility

Our findings revealed that 22 studies failed to report and discuss the NPV, PPV, NNS, NNR, NRI, and net benefit or decision curve (see Table 2 for the definitions of these metrics). Fifteen studies adopted arbitrary cutoffs for risk threshold determination, and ten did not propose plausible risk thresholds (Table 5). Furthermore, only a few studies reported the sensitivity (recall or true positive rate; $n = 16$) and specificity (true negative rate; $n = 15$) (Table 4). Of note, only five studies demonstrated high potential clinical utility for their models. These studies not only reported model performance but also employed appropriate risk threshold determination, exhibited clinical usefulness to aid in decision making, discussed standard metrics understandable to clinicians (e.g., PPV and NPV), and reported incremental measures such as the NRI. Specifically, Sutherland et al. (2021) reported the model performance, presented multiple plausible risk thresholds based on predicted probabilities, and reported the missed rate for high-risk adenomas (HRAs) of 5% risk threshold [40]. They demonstrated the clinical utility of their model by aiming at identifying individuals with HRAs during a primary screening colonoscopy. They used multiple factors, such as age, sex, and BMI, to develop this model, which achieved an optimism-adjusted AUC of 67%. The practical application of the model was further highlighted when they indicated that, by adopting a 5% risk threshold, the model could potentially eliminate the need for an immediate colonoscopy in about a third of the tested individuals. Even with a 13% miss rate for HRAs at this threshold, these could still be detected through subsequent FIT screenings and suggested that the model could aid clinicians in prioritizing individuals for screening colonoscopy.

Furthermore, among these five studies, only one discussed the net benefit of the model, a critical metric enabling clinicians to balance the advantages of true positives against the disadvantages of false positives for a given threshold of risks. This net benefit was demonstrated by Briggs et al. (2023), who presented and discussed a decision curve analysis [43]. They assessed the clinical utility of the model (QCancer-10 model + polygenic risk score (PRS)) by proposing a scenario in which one colonoscopy at the onset of an eight-year span detects all colorectal anomalies. Using a net benefit formula that factors in true positives, false positives, and risk thresholds, the findings were visualized with decision curves. Their finding showed that a small incremental improvement in net benefit with QCancer-10 plus polygenic risk score models compared with QCancer-10 alone was observed, implying that integrating the PRS with the QCancer-10 model has a modest enhancement in the overall benefit of predictive CRC risk when compared to using only the QCancer-10 model [43].

3.9.3. Clinical Usability

In terms of clinical usability, only eight studies reported the use of a simple risk score or assessment tool to estimate the risk of CRC [38,45,46,49,63,65–67]. Meanwhile, 13 studies reported a clear risk threshold that only used less than four tiers to facilitate risk-stratification of individuals [40,41,47,50,51,53,54,57,60,62,68,70,71] (Supplementary Materials Figure S1).

4. Discussion

4.1. Summary Findings

Previous systematic reviews have evaluated risk prediction models for CRC or AN, including a head-to-head comparison and a meta-analysis of the model performance of risk scores based on cohorts of average risk or asymptomatic populations undergoing colonoscopy [8,23]. These reviews indicated that risk scores exhibited moderate discriminatory accuracy and suggested integrating advanced features, like novel laboratory findings or polygenic risk scores, to enhance risk score performance.

This review included a comprehensive synthesis and appraisal of 41 studies (44 models), including 26 original conventional models and 12 composite models. Of these, 12 conventional and 4 composite models underwent both internal and external validations, while 21 de novo models were limited to internally validation and four did not undergo any form of validation. All models reported varying discriminatory accuracy, with AUCs ranging

from 0.57 (conventional model) to 0.90 (composite model). However, all except the study of van 't Klooster et al. exhibited an overall unclear or high risk of bias, and 12 raised concerns about applicability [76]. Pronounced biases included the exclusion of relevant participants, lack of blinding, improper statistical methods, and small sample sizes.

Among the five models that demonstrated high potential for clinical utility (⊕⊕⊕⊕), only three underwent external validation, including one conventional model and two composite models. Both models demonstrated modest discriminatory accuracy and good calibration. However, only one model has shown high potential for clinical utility with an overall low risk of bias, albeit with limited generalizability since it exclusively considered individuals with a well-established cardiovascular disease. Of note, only one of these studies elaborated on the net benefit of the model [43]. The authors introduced a scenario leveraging the QCancer-10 model combined with a polygenetic risk score (PRS) and showcased an incremental improvement when integrating a PRS, enhancing the overall benefit of the model for predictive CRC risk assessment. Nonetheless, the study by Sutherland et al. (2021) also well demonstrated the clinical utility of their model. They comprehensively elaborated on the presented performance of the model, as well as its capacity to inform and refine clinical decision making, such as reducing immediate procedural needs for screening colonoscopy [40]. They also presented a scenario based on a given risk threshold and emphasized complementary approaches like FIT screenings to address the missed rate of HRAs.

In terms of clinical usability, the APCS scores, validated externally in four separate studies, raised concerns about their potential clinical utility. Nevertheless, they demonstrated clinical usability through a simple risk score and received recent recommendations for adoption from the Asia-Pacific Working Group. Our study highlights the need for further external validation, unbiased approaches to model development and validation, and enhanced reporting of the potential clinical utility and usability of the model.

We further explored the characteristics of the externally validated models with high potential for clinical utility. These models include two composite models developed and externally validated by Thomsen et al. (2022) and van 't Klooster et al. (2020), and Kaminski's risk score, a conventional model, which was externally validated by Ruco et al. (2015). One of the highlighted models was the risk prediction model developed by Thomsen et al., which not only showed promise for clinical utility but also demonstrated high discriminatory accuracy, good calibration, and high clinical usability [41]. Interestingly, their model involved only a combination of a FIT test result and sociodemographic factors—age and sex. This is because their study focused on developing a prediction model for risk-stratified screening, specifically for the purpose of selecting participants for diagnostic colonoscopy after FIT-based CRC screening. The proposed predictive models effectively estimated the risk of CRC or ACN by using age, gender, and fecal hemoglobin levels, which are readily available in all FIT-based screening protocols [41]. Conversely, van 't Klooster's model exhibited modest discriminatory accuracy and good calibration. However, its generalizability was limited due to its exclusive focus on predicting the risk of CRC among patients with cardiovascular disease. Furthermore, Kaminski's validated model displayed high potential for clinical utility and generalizability [60]; however, its complex risk stratification approach posed challenges, thus limiting the model's usability. More precisely, despite successfully establishing risk cutoff points, it demonstrated complexity due to risk stratification involving up to eight tiers, which may pose difficulties in decision making for clinicians and individuals considering a screening colonoscopy.

Several models have also been externally validated but exhibited limited clinical utility due to a lack of guidance on their practical implementation in clinical settings. Specifically, the studies conducted by Cai et al. (2012) [37], Shin et al. (2014) [48], Sekiguchi et al. (2018) [59], Sharara et al. (2020) [58], Wei et al. (2017) [78], and Yen et al. (2014) [56] revealed limited clinical utility despite their noteworthy discriminatory accuracy. This is attributed to instances where these studies presented risk stratification methodologies without well-defined cutoffs or did not report plausible risk stratification approaches.

4.2. Common Pitfalls of CRC Risk Prediction Models

Our review findings revealed that several risk models, including both conventional and composite models, demonstrated modest-to-high discriminatory accuracy and good calibration. However, less than half of these models exhibited clinical usability, and their generalizability was limited. These limitations stemmed from various pitfalls in developing, validating, and reporting risk prediction models, which were identified when assessing the risk of bias, clinical usability, potential clinical utility, and generalizability of the models:

- Exclusion of relevant populations: Key populations, such as those with diabetes or cardiovascular diseases, were often excluded. Excluding participants with a family history of CRC is particularly concerning given its established role as a significant risk factor.
- Non-accounting for competing risk/censoring: Failure to account for competing risks or censoring events (e.g., deaths or withdrawals) can distort risk estimates, compromising model accuracy and reliability.
- Non-blinding of predictor and outcome assessors: In certain studies, assessors lacked blinding, potentially introducing bias, as they might unconsciously favor certain predictors or outcomes, thus affecting model development and evaluation.
- Reliance on univariate analysis for risk predictor selection: In studies that exclusively relied on univariate analysis for predictor selection, a potential drawback emerged wherein significant predictors might be inadvertently omitted from consideration. This risk stems from the failure to account for confounding bias, as a univariate analysis examines predictors in isolation, neglecting potential interactions with other variables. Consequently, influential predictors could be excluded from the analysis due to the presence of these confounding pathways.
- Use of split-sample techniques in conducting model validation in most studies: The limitations of the split-sampling method include sensitivity to data split, limited training data, sampling variability, assumption of constant data distribution, and the potential for overfitting or underfitting. These issues were less pronounced in alternative methods like bootstrapping or 10-fold cross-validation, which were employed by only nine studies included in this review.
- Lack of information regarding the model's stability: Most studies did not report E/O ratios and calibration plots, thus hindering our understanding of the model's stability across different scenarios and the distribution of risk estimates among all individuals in a dataset.
- Unclear cutoffs or the use of arbitrary risk threshold determination: Inappropriate determination of risk thresholds, characterized by unclear cutoffs or arbitrary choices, can undermine the clinical utility of the models. These issues may also lead to misclassification of risk, reduced sensitivity or specificity, lack of standardization, and difficulties in patient communication.
- Low sample size when conducting external validation and lack of external validation of most models. Numerous risk models developed to estimate CRC exist, but less than half of these new models have been externally validated and rarely had demonstrated potential clinical utility. Moreover, several included models in this review often suffer from small sample sizes during external validations. Furthermore, many of the models examined in this review frequently encountered the issue of insufficient sample sizes, especially when undergoing external validation. Small sample sizes can result in imprecise estimates of a model's performance metrics, such as calibration, discrimination, and clinical utility [81].
- Inefficient handling of continuous variables: Some studies demonstrated poor handling of continuous variables like age and BMI without proper justification, which was observed in most of the included studies. Research has demonstrated that categorizing continuous predictors results in models that exhibit poor predictive performance and limited clinical utility. The act of categorizing continuous predictors is both unneces-

sary and biologically implausible, proving to be an inefficient practice that should be avoided in the development of risk models [30].

- Non-reporting of net benefits or decision curves. None of the studies demonstrated clinical scenarios or decision nodes, such as net benefits and decision curves, associated with risk categories or multiple plausible thresholds.

These observations underscore the need for more rigorous and unbiased approaches when developing and validating CRC risk prediction models, as well as the importance of standardized reporting that accounts not only for the model performance but also for their clinical usability, clinical utility, and generalizability of the model.

4.3. Recommendation and Implications for Research

We make propositions for model development, validation, and enhanced reporting with regard to potential clinical utility and usability of the risk model.

Proposition 1. *The exclusion of individuals with a family history of CRC during the development of risk prediction models warrants careful consideration, given its well-established status as a risk factor for CRC [82]. Clear reasons for such exclusion and transparency during consent are essential. Therefore, it is imperative to ensure that these individuals are not only informed about their exclusion but also guided towards appropriate screening and preventive measures that are in line with current guidelines. Additionally, excluding those with comorbidities can introduce selection bias, potentially limiting the generalizability of study findings. Moreover, prediction models should accurately handle competing events, using methods that prevent bias in risk predictions, as this might go unnoticed in external validations [83].*

Proposition 2. *The following recommendations for determining risk thresholds to enhance the clinical utility of risk prediction models are firmly grounded in the principles of PROGRESS (Prognosis Research Strategy) 4 [15,84] and were supported by the research of Wynants et al. in 2019 [30]. In accordance with this guideline, the emphasis lies in aligning risk thresholds with the clinical implications of risk stratification, striving for levels of risk considered acceptable within the specific clinical context [84]. Furthermore, the recommendation is to present multiple plausible risk thresholds to cater to diverse clinical scenarios, thereby enhancing the overall utility of the model [30]. These guidelines emphasize that plausible risk thresholds should be based on clinical or theoretical grounds rather than just metrics like Youden's index, which might be insufficient in clinical settings. It is important to consider factors like intervention costs and the relevance of colonoscopy, especially with well-calibrated models. For models not perfectly calibrated, it is recommended to consider the population prevalence of colorectal cancer, as well as the sensitivity and specificity of the model, when determining risk thresholds [30,84].*

Proposition 3. *To address the issue on low sample sizes and determine the minimum sample size required for a new external validation study, the proposed calculations consider certain factors, including the desired level of precision (confidence interval width), the expected event proportion in the validation population, the expected (mis)calibration of the model, the variance of predictor values, and potential risk thresholds for clinical decision making [81].*

In studies where models or risk scores with high clinical usability were identified, we underscored a common approach characterized by the development of a simple risk scoring system. As an exemplar of this approach, the (modified) APCS scores have demonstrated high clinical usability due to their simple risk scoring system, which makes it easy to implement them in practice and effectively stratify individuals into risk categories for CRC screening in the Asia–Pacific region [11,45,49]. Specifically, the model was presented using a risk score ranging from 0 to 7 points, stratifying patients into three risk tiers: scores 0 to 1 were categorized as low risk (LR), scores 2 to 3 as moderate risk (MR), and scores 4 to 7 as high risk (HR). The risk score was calculated as the sum of individual risk factors, whether present or absent in an individual. The simplicity of APCS risk score was attributed to its

methodology of assigning weights to individual variables based on their adjusted odds ratios (AORs) [45,49]. To achieve this simplicity, the AOR values were halved and then rounded to the nearest integer, effectively constraining the total risk score for each subject to remain below 10 [49]. This approach was implemented to streamline the scoring system, making it more straightforward and practical for clinical application.

Proposition 4. *In the context of handling continuous predictors, an alternative approach was proposed, wherein clinical decisions should be founded upon the estimated risk, rather than categorizing predictors as dichotomous or categorical variables [30]. If risk groups are desired, these should be defined based on a predicted outcome of the model instead of inputs that are recommended by these guidelines. However, our research uncovered a noteworthy study conducted by Liu et al. (2018), who conducted a comparative analysis between a concise categorized lifestyle exposure-based colorectal cancer (CRC) risk prediction tool and a model employing continuous measures [67].*

It is intriguing that the results indicated that the categorical assessment of predictor variables did not result in a loss of model performance in the general population. One possible explanation for the insignificant performance difference between the categorized lifestyle exposure-based CRC risk prediction tool and the model using continuous measures could be related to the specific dataset, population, and, most importantly, the risk factor under investigation. It is possible that the chosen lifestyle exposure variables in their study naturally lent themselves to categorization or that the population studied had characteristics that made categorization effective.

Additionally, the modeling techniques and algorithms employed may have contributed to this result. To address these conflicting findings, it is important to consider the broader context and conduct further research. A comprehensive meta-analysis or systematic review of existing studies in this domain may help identify patterns and factors that influence the choice between categorical and continuous predictors. Nevertheless, the choice between categorical and continuous predictors should be guided by their practical implication and clinical utility—in this case, in the context of a CRC risk-stratified screening.

Proposition 5. *Model stability, as evidenced through calibration plots, is crucial. Consider a model developed in the US and later used in Belgium. It is essential to verify if the model performs consistently in the new context. If not, calibration using local data is necessary to adjust its predictions. Furthermore, as time progresses, models initially designed years prior should be periodically reviewed and recalibrated if they no longer align with current standards to ensure ongoing accuracy and relevance.*

Proposition 6. *Researchers engaged in risk prediction modeling should give precedence to the decision curve analysis for assessing the clinical utility of their models. This approach comprehensively integrates both the benefits and limitations of the prediction model in comparison with a default strategy [85]. The decision curves analysis enables us to have a better understanding of the consequences of the decisions made that are driven by a model or test. This is a perspective that traditional metrics such as discrimination and calibration often miss. To further learn how to conduct a decision curve analysis, studies by Vickers et al. (2006 and 2019) on the decision curve analysis are recommended [85,86]. Additionally, considering the widespread misunderstanding of and confusion about the decision curve analysis, they further presented a step-by-step didactic introduction to interpreting a decision curve analysis.*

5. Conclusions

In this review, certain risk models for CRC were highlighted that display a promising performance, clinical utility, clinical usability, and high generalizability. Interestingly, one model has been adopted in the Asia–Pacific region due to its simple risk scoring system, particularly when combined with FIT results. Unfortunately, numerous risk models for CRC still lack external validation and have low potential for clinical utility due to the non-

reporting of interpretable metrics, inadequate assessment of model stability, inappropriate statistical approaches, poor handling of continuous predictors, and unclear or improper determination of risk thresholds. Moreover, a significant gap in the current research landscape is the absence of a decision curve analysis, a crucial tool to understand the net benefits and implications of these models compared with a default strategy in actual clinical settings. It is important to emphasize that a risk model with a moderate performance but high potential for clinical utility can be more impactful than a high-performing model lacking real-world relevance. In essence, a real-world application of the model and impact on patient care define its true value, emphasizing the importance of clinical utility over mere accuracy.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/gidisord5040045/s1>.

Author Contributions: D.J.H. conceptualized and developed the review protocol, conducted abstract and full-text screenings for all articles, developed data extraction forms, extracted data, assessed the risk of bias for all included studies, validated all information extracted and assessed by other authors, drafted the initial manuscript, and revised and finalized the manuscript write-up. G.V.H. contributed to the conceptualization, retrieved hard-to-access full-text literature, reviewed the protocol and original draft, provided validation, reviewed the draft and final manuscript, and offered supervision. W.v.d.V. was involved in the conceptualization, abstract, and full-text screening independent of D.J.H.; comprehensively reviewed the original draft and final manuscript revision; and provided supervision. D.M.S. conducted data validation, provided guidance on data extraction related to model performance and other critical metrics, independently assessed the risk of bias for included studies, and reviewed both the original draft and the final manuscript. M.M.B., C.G.-O., and N.B.Y. conducted data extraction, independently assessed the risk of bias of the selected studies, and reviewed the original and final draft. K.F. validated extracted data, assessed the risk of bias for included studies, and reviewed the original and final draft. A.F. reviewed the protocol and conducted abstract screening of articles. All authors have read and agreed to the published version of the manuscript.

Funding: This study is part of a prevention project realized with the support of the Fight Against Cancer project (Kom op Tegen Kanker) (Funding number: 800300025). The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. All authors had full access to all the data in the study and had the final responsibility for the decision to submit for publication.

Data Availability Statement: The data extracted and synthesized for this systematic review can be found in the Supplementary Materials. Due to the complexity and volume of certain data, such as the R codes used for radial chart creation and comprehensive data extraction forms, these detailed materials are available upon request.

Acknowledgments: We are immensely thankful to the ORIENT Research team for their guidance and unwavering support throughout this review.

Conflicts of Interest: The authors declare no competing interests, and all authors' contributions are duly documented.

References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [[CrossRef](#)]
2. Keum, N.; Giovannucci, E. Global burden of colorectal cancer: Emerging trends, risk factors and prevention strategies. *Nat. Rev. Gastroenterol. Hepatol.* **2019**, *16*, 713–732. [[CrossRef](#)] [[PubMed](#)]
3. Edwards, B.K.; Ward, E.; Kohler, B.A.; Ehemann, C.; Zaubler, A.G.; Anderson, R.N.; Jemal, A.; Schymura, M.J.; Lansdorp-Vogelaar, I.; Seeff, L.C.; et al. Annual report to the nation on the status of cancer, 1975–2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. *Cancer* **2010**, *116*, 544–573. [[CrossRef](#)] [[PubMed](#)]
4. Arnold, M.; Sierra, M.S.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global patterns and trends in colorectal cancer incidence and mortality. *Gut* **2017**, *66*, 683–691. [[CrossRef](#)]
5. McGeoch, L.; Saunders, C.L.; Griffin, S.J.; Emery, J.D.; Walter, F.M.; Thompson, D.J.; Antoniou, A.C.; Usher-Smith, J.A. Risk Prediction Models for Colorectal Cancer Incorporating Common Genetic Variants: A Systematic Review. *Cancer Epidemiol. Biomark. Prev.* **2019**, *28*, 1580–1593. [[CrossRef](#)]

6. Lansdorp-Vogelaar, I.; Meester, R.; de Jonge, L.; Buron, A.; Haug, U.; Senore, C. Risk-stratified strategies in population screening for colorectal cancer. *Int. J. Cancer* **2021**, *150*, 397–405. [[CrossRef](#)] [[PubMed](#)]
7. Ponti, A.; Anttila, A.; Ronco, G.; Senore, C. Cancer Screening in the European Union (2017). International Agency for Research on Cancer. France. 2017. Available online: https://health.ec.europa.eu/system/files/2017-05/2017_cancerscreening_2ndreportimplementation_en_0.pdf (accessed on 6 November 2023).
8. Peng, L.; Balavarca, Y.; Weigl, K.; Hoffmeister, M.; Brenner, H. Head-to-Head Comparison of the Performance of 17 Risk Models for Predicting Presence of Advanced Neoplasms in Colorectal Cancer Screening. *Am. J. Gastroenterol.* **2019**, *114*, 1520–1530. [[CrossRef](#)] [[PubMed](#)]
9. World Health Organization. A Short Guide to Cancer Screening Increase Effectiveness, Maximize Benefits and Minimize Harm. Co-Penhagen: WHO Regional Office for Europe. 2022. Available online: <http://apps.who.int/bookorders> (accessed on 17 October 2023).
10. Hull, M.A.; Rees, C.J.; Sharp, L.; Koo, S. A risk-stratified approach to colorectal cancer prevention and diagnosis. *Nat. Rev. Gastroenterol. Hepatol.* **2020**, *17*, 773–780. [[CrossRef](#)]
11. Sung, J.J.Y.; Chiu, H.-M.; Lieberman, D.; Kuipers, E.J.; Rutter, M.D.; Macrae, F.; Yeoh, K.-G.; Ang, T.L.; Chong, V.H.; John, S.; et al. Third Asia-Pacific consensus recommendations on colorectal cancer screening and postpolypectomy surveillance. *Gut* **2022**, *71*, 2152–2166. [[CrossRef](#)]
12. Herrera, D.J.; van de Veerdonk, W.; Berhe, N.M.; Talboom, S.; van Loo, M.; Alejos, A.R.; Ferrari, A.; Van Hal, G. Mixed-Method Systematic Review and Meta-Analysis of Shared Decision-Making Tools for Cancer Screening. *Cancers* **2023**, *15*, 3867. [[CrossRef](#)]
13. Debray, T.P.A.; Damen, J.A.A.G.; Snell, K.I.E.; Ensor, J.; Hooft, L.; Reitsma, J.B.; Riley, R.D.; Moons, K.G.M. A guide to systematic review and meta-analysis of prediction model performance. *BMJ* **2017**, *356*, i6460. [[CrossRef](#)] [[PubMed](#)]
14. Usher-Smith, J.A.; Walter, F.M.; Emery, J.D.; Win, A.K.; Griffin, S.J. Risk Prediction Models for Colorectal Cancer: A Systematic Review. *Cancer Prev. Res.* **2016**, *9*, 13–26. [[CrossRef](#)]
15. Riley, R.D.; Hayden, J.A.; Steyerberg, E.W.; Moons, K.G.M.; Abrams, K.; Kyzas, P.A.; Malats, N.; Briggs, A.; Schroter, S.; Altman, D.G.; et al. Prognosis Research Strategy (PROGRESS) 2: Prognostic Factor Research. *PLoS Med.* **2013**, *10*, e1001380. [[CrossRef](#)] [[PubMed](#)]
16. Zheng, Y.; Hua, X.; Win, A.K.; MacInnis, R.J.; Gallinger, S.; Le Marchand, L.; Lindor, N.M.; Baron, J.A.; Hopper, J.L.; Dowty, J.G.; et al. A New Comprehensive Colorectal Cancer Risk Prediction Model Incorporating Family History, Personal Characteristics, and Environmental Factors. *Cancer Epidemiol. Biomark. Prev.* **2020**, *29*, 549–557. [[CrossRef](#)] [[PubMed](#)]
17. Matthias, M.S.; Imperiale, T.F. A risk prediction tool for colorectal cancer screening: A qualitative study of patient and provider facilitators and barriers. *BMC Fam. Pr.* **2020**, *21*, 43. [[CrossRef](#)]
18. Li, R.; Duan, R.; He, L.; Moore, J.H. Risk Prediction: Methods, Challenges, and Opportunities. 2022. Available online: www.worldscientific.com (accessed on 17 October 2023).
19. Luo, J.-C.; Zhao, Q.-Y.; Tu, G.-W. Clinical prediction models in the precision medicine era: Old and new algorithms. *Ann. Transl. Med.* **2020**, *8*, 274. [[CrossRef](#)]
20. Saya, S.; Emery, J.D.; Dowty, J.G.; McIntosh, J.G.; Winship, I.M.; Jenkins, M.A. The Impact of a Comprehensive Risk Prediction Model for Colorectal Cancer on a Population Screening Program. *JNCI Cancer Spectr.* **2020**, *4*, pkaa062. [[CrossRef](#)] [[PubMed](#)]
21. Stiglic, G.; Kocbek, P.; Fijacko, N.; Zitnik, M.; Verbert, K.; Cilar, L. Interpretability of machine learning-based prediction models in healthcare. *WIREs Data Min. Knowl. Discov.* **2020**, *10*, e1379. [[CrossRef](#)]
22. Tangri, N.; Kitsios, G.D.; Inker, L.A.; Griffith, J.; Naimark, D.M.; Walker, S.; Rigatto, C.; Uhlig, K.; Kent, D.M.; Levey, A.S. Risk Prediction Models for Patients with Chronic Kidney Disease A Systematic Review. 2013. Available online: www.annals.org (accessed on 17 October 2023).
23. Peng, L.; Weigl, K.; Boakye, D.; Brenner, H. Risk Scores for Predicting Advanced Colorectal Neoplasia in the Average-risk Population: A Systematic Review and Meta-analysis. *Am. J. Gastroenterol.* **2018**, *113*, 1788–1800. [[CrossRef](#)]
24. Damen, J.A.; Moons, K.G.; van Smeden, M.; Hooft, L. How to conduct a systematic review and meta-analysis of prognostic model studies. *Clin. Microbiol. Infect.* **2022**, *29*, 434–440. [[CrossRef](#)]
25. Wolff, R.; Moons, K.; Riley, R.; Whiting, P.F.; Westwood, M.; Collins, G.S. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann. Intern. Med.* **2019**, *1*, 51–58. [[CrossRef](#)]
26. Higgins, J.P.T.; Thomas, J.; Chandler, J.; Cumpston, M.; Li, T.; Page, M.J.; Welch, V.A. *Cochrane Handbook for Systematic Reviews of Interventions*; John Wiley & Sons: Hoboken, NJ, USA, 2019; Available online: <https://training.cochrane.org/handbook> (accessed on 6 November 2023).
27. Moons, K.G.; Wolff, R.F.; Riley, R.D.; Whiting, P.F.; Westwood, M.; Collins, G.S.; Reitsma, J.B.; Kleijnen, J.; Mallett, S. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Ann. Intern. Med.* **2019**, *1*, W1–W33. [[CrossRef](#)]
28. Kreuzberger, N.; Damen, J.A.; Trivella, M.; Estcourt, L.J.; Aldin, A.; Umlauff, L.; Vazquez-Montes, M.D.; Wolff, R.; Moons, K.G.; Monsef, I.; et al. Prognostic models for newly-diagnosed chronic lymphocytic leukaemia in adults: A systematic review and meta-analysis. *Cochrane Database Syst. Rev.* **2020**, *7*, CD012022. [[CrossRef](#)]
29. Debray, T.P.; Damen, J.A.; Riley, R.D.; Snell, K.; Reitsma, J.B.; Hooft, L.; Collins, G.S.; Moons, K.G. A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. *Stat. Methods Med. Res.* **2019**, *28*, 2768–2786. [[CrossRef](#)]

30. Wynants, L.; Van Smeden, M.; McLernon, D.J.; Timmerman, D.; Steyerberg, E.W.; Van Calster, B. Three myths about risk thresholds for prediction models. *BMC Med.* **2019**, *17*, 192. [[CrossRef](#)] [[PubMed](#)]
31. Louro, J.; Posso, M.; Boon, M.H.; Román, M.; Domingo, L.; Castells, X.; Sala, M. A systematic review and quality assessment of individualised breast cancer risk prediction models. *Br. J. Cancer* **2019**, *121*, 76–85. [[CrossRef](#)]
32. Kerr, K.F.; Wang, Z.; Janes, H.; McClelland, R.L.; Psaty, B.M.; Pepe, M.S. Net Reclassification Indices for Evaluating Risk Prediction Instruments. *Epidemiology* **2014**, *25*, 114–121. [[CrossRef](#)] [[PubMed](#)]
33. Janssens, A.C.J.W.; Martens, F.K. Reflection on modern methods: Revisiting the area under the ROC Curve. *Leuk. Res.* **2020**, *49*, 1397–1403. [[CrossRef](#)] [[PubMed](#)]
34. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022.
35. van de Veerdonk, W.; Hoeck, S.; Peeters, M.; Van Hal, G. Towards risk-stratified colorectal cancer screening. Adding risk factors to the fecal immunochemical test: Evidence, evolution and expectations. *Prev. Med.* **2019**, *126*, 105746. [[CrossRef](#)]
36. Brand, E.C.; Crook, J.E.; Thomas, C.S.; Siersema, P.D.; Rex, D.K.; Wallace, M.B. Development and validation of a prediction model for adenoma detection during screening and surveillance colonoscopy with comparison to actual adenoma detection rates. *PLoS ONE* **2017**, *12*, e0185560. [[CrossRef](#)]
37. Cai, Q.-C.; Yu, E.-D.; Xiao, Y.; Bai, W.-Y.; Chen, X.; He, L.-P.; Yang, Y.-X.; Zhou, P.-H.; Jiang, X.-L.; Xu, H.-M.; 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. [[CrossRef](#)] [[PubMed](#)]
38. Cao, Y.; Rosner, B.A.; Ma, J.; Tamimi, R.M.; Chan, A.T.; Fuchs, C.S.; Wu, K.; Giovannucci, E.L. Assessing individual risk for high-risk colorectal adenoma at first-time screening colonoscopy. *Int. J. Cancer* **2015**, *137*, 1719–1728. [[CrossRef](#)] [[PubMed](#)]
39. Stegeman, I.; de Wijckerslooth, T.R.; Stoop, E.M.; van Leerdam, M.E.; Dekker, E.; van Ballegooijen, M.; Kuipers, E.J.; Fockens, P.; Kraaijenhagen, R.A.; Bossuyt, P.M.; et al. Combining risk factors with faecal immunochemical test outcome for selecting CRC screenees for colonoscopy. *Gut* **2014**, *63*, 466–471. [[CrossRef](#)] [[PubMed](#)]
40. Sutherland, R.L.; Boyne, D.J.; Jarada, T.N.; Lix, L.M.; Timmouth, J.; Rabeneck, L.; Heitman, S.J.; Forbes, N.; Hilsden, R.J.; Brenner, D.R. Development and validation of a risk prediction model for high-risk adenomas at the time of first screening colonoscopy among screening aged Canadians. *Prev. Med.* **2021**, *148*, 106563. [[CrossRef](#)]
41. Thomsen, M.K.; Pedersen, L.; Erichsen, R.; Lash, T.L.; Sørensen, H.T.; Mikkelsen, E.M. Risk-stratified selection to colonoscopy in FIT colorectal cancer screening: Development and temporal validation of a prediction model. *Br. J. Cancer* **2022**, *126*, 1229–1235. [[CrossRef](#)] [[PubMed](#)]
42. Meester, R.G.S.; van de Schootbrugge-Vandermeer, H.J.; Breekveldt, E.C.H.; de Jonge, L.; Toes-Zoutendijk, E.; Kooyker, A.; Nieboer, D.; Ramakers, C.R.; Spaander, M.C.W.; van Vuuren, A.J.; et al. Faecal occult blood loss accurately predicts future detection of colorectal cancer. A prognostic model. *Gut* **2023**, *72*, 101–108. [[CrossRef](#)]
43. Briggs, S.E.W.; Law, P.; East, J.E.; Wordsworth, S.; Dunlop, M.; Houlston, R.; Hippisley-Cox, J.; Tomlinson, I. Integrating genome-wide polygenic risk scores and non-genetic risk to predict colorectal cancer diagnosis using UK Biobank data: Population based cohort study. *BMJ* **2022**, *379*, e071707. [[CrossRef](#)]
44. Müllder, D.T.; Puttelaar, R.V.D.; Meester, R.G.; O'Mahony, J.F.; Lansdorp-Vogelaar, I. Development and validation of colorectal cancer risk prediction tools: A comparison of models. *Int. J. Med. Inform.* **2023**, *178*, 105194. [[CrossRef](#)]
45. Yeoh, K.-G.; Ho, K.-Y.; Chiu, H.-M.; Zhu, F.; Ching, J.Y.L.; Wu, D.-C.; Matsuda, T.; Byeon, J.-S.; Lee, S.-K.; Goh, K.-L.; 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. [[CrossRef](#)]
46. He, X.-X.; Yuan, S.-Y.; Li, W.-B.; Yang, H.; Ji, W.; Wang, Z.-Q.; Hao, J.-Y.; Chen, C.; Chen, W.-Q.; Gao, Y.-X.; et al. Improvement of Asia-Pacific colorectal screening score and evaluation of its use combined with fecal immunochemical test. *BMC Gastroenterol.* **2019**, *19*, 226. [[CrossRef](#)]
47. Yang, H.-J.; Choi, S.; Park, S.-K.; Jung, Y.S.; Choi, K.Y.; Park, T.; Kim, J.Y.; Park, D.I. Derivation and validation of a risk scoring model to predict advanced colorectal neoplasm in adults of all ages. *J. Gastroenterol. Hepatol.* **2017**, *32*, 1328–1335. [[CrossRef](#)] [[PubMed](#)]
48. Shin, A.; Joo, J.; Yang, H.-R.; Bak, J.; Park, Y.; Kim, J.; Oh, J.H.; Nam, B.-H. Risk Prediction Model for Colorectal Cancer: National Health Insurance Corporation Study, Korea. *PLoS ONE* **2014**, *9*, e88079. [[CrossRef](#)] [[PubMed](#)]
49. Sung, J.J.Y.; Wong, M.C.S.; Lam, T.Y.T.; Tsoi, K.K.F.; Chan, V.C.W.; Cheung, W.; Ching, J.Y.L. A modified colorectal screening score for prediction of advanced neoplasia: A prospective study of 5744 subjects. *J. Gastroenterol. Hepatol.* **2018**, *33*, 187–194. [[CrossRef](#)] [[PubMed](#)]
50. Kim, J.Y.; Choi, S.; Park, T.; Kim, S.K.; Jung, Y.S.; Park, J.H.; Kim, H.J.; Cho, Y.K.; Sohn, C.I.; Jeon, W.K.; et al. Development and validation of a scoring system for advanced colorectal neoplasm in young Korean subjects less than age 50 years. *Intest. Res.* **2019**, *17*, 253–264. [[CrossRef](#)] [[PubMed](#)]
51. Hong, S.N.; Son, H.J.; Choi, S.K.; Chang, D.K.; Kim, Y.-H.; Jung, S.-H.; Rhee, P.-L. A prediction model for advanced colorectal neoplasia in an asymptomatic screening population. *PLoS ONE* **2017**, *12*, e0181040. [[CrossRef](#)]
52. Imperiale, T.F.; Monahan, P.O.; Stump, T.E.; Ransohoff, D.F. Derivation and validation of a predictive model for advanced colorectal neoplasia in asymptomatic adults. *Gut* **2021**, *70*, 1155–1161. [[CrossRef](#)] [[PubMed](#)]
53. Imperiale, T.F.; Monahan, P.O.; Stump, T.E.; Glowinski, E.A.; Ransohoff, D.F. Derivation and Validation of a Scoring System to Stratify Risk for Advanced Colorectal Neoplasia in Asymptomatic Adults. *Ann. Intern. Med.* **2015**, *163*, 339–346. [[CrossRef](#)]

54. Schroy, P.C.; Wong, J.B.; O'Brien, M.J.; Chen, C.A.; Griffith, J.L. A Risk Prediction Index for Advanced Colorectal Neoplasia at Screening Colonoscopy. *Am. J. Gastroenterol.* **2015**, *110*, 1062–1071. [[CrossRef](#)]
55. 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. [[CrossRef](#)]
56. Yen, A.M.; Chen, S.L.; Chiu, S.Y.; Fann, J.C.; Wang, P.; Lin, S.; Chen, Y.; Liao, C.; Yeh, Y.; Lee, Y.; et al. A new insight into fecal hemoglobin concentration-dependent predictor for colorectal neoplasia. *Int. J. Cancer* **2014**, *135*, 1203–1212. [[CrossRef](#)]
57. Soonklang, K.M.; Siribumrungwong, B.; Siripongpreeda, B.; Auewarakul, C. Comparison of multiple statistical models for the development of clinical prediction scores to detect advanced colorectal neoplasms in asymptomatic Thai patients. *Medicine* **2021**, *100*, e26065. [[CrossRef](#)]
58. Sharara, A.I.; El Mokahal, A.; Harb, A.H.; Khalaf, N.; Sarkis, F.S.; El-Halabi, M.M.; Mansour, N.M.; Malli, A.; Habib, R. Risk prediction rule for advanced neoplasia on screening colonoscopy for average-risk individuals. *World J. Gastroenterol.* **2020**, *26*, 5705–5717. [[CrossRef](#)]
59. Sekiguchi, M.; Kakugawa, Y.; Matsumoto, M.; Matsuda, T. A scoring model for predicting advanced colorectal neoplasia in a screened population of asymptomatic Japanese individuals. *J. Gastroenterol.* **2018**, *53*, 1109–1119. [[CrossRef](#)] [[PubMed](#)]
60. Ruco, A.; Stock, D.; Hilsden, R.J.; McGregor, S.E.; Paszat, L.F.; Saskin, R.; Rabeneck, L. Evaluation of a clinical risk index for advanced colorectal neoplasia among a North American population of screening age. *BMC Gastroenterol.* **2015**, *15*, 162. [[CrossRef](#)] [[PubMed](#)]
61. Jung, Y.S.; Park, C.H.; Kim, N.H.; Park, J.H.; Park, D.I.; Sohn, C.I. Clinical risk stratification model for advanced colorectal neoplasia in persons with negative fecal immunochemical test results. *PLoS ONE* **2018**, *13*, e0191125. [[CrossRef](#)] [[PubMed](#)]
62. Park, C.H.; Jung, Y.S.; Kim, N.H.; Park, J.H.; Park, D.I.; Sohn, C.I. Usefulness of risk stratification models for colorectal cancer based on fecal hemoglobin concentration and clinical risk factors. *Gastrointest. Endosc.* **2019**, *89*, 1204–1211.e1. [[CrossRef](#)] [[PubMed](#)]
63. Musselwhite, L.W.; Redding, T.S.; Sims, K.J.; O'leary, M.C.; Hauser, E.R.; Hyslop, T.; Gellad, Z.F.; Sullivan, B.A.; Lieberman, D.; Provenzale, D. Advanced neoplasia in Veterans at screening colonoscopy using the National Cancer Institute Risk Assessment Tool. *BMC Cancer* **2019**, *19*, 1097. [[CrossRef](#)]
64. Murchie, B.; Tandon, K.; Hakim, S.; Shah, K.; O'rourke, C.; Castro, F.J. A New Scoring System to Predict the Risk for High-risk Adenoma and Comparison of Existing Risk Calculators. *J. Clin. Gastroenterol.* **2017**, *51*, 345–351. [[CrossRef](#)]
65. 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*, 534–541. [[CrossRef](#)]
66. Luu, X.Q.; Lee, K.; Kim, J.; Sohn, D.K.; Shin, A.; Choi, K.S. The classification capability of the Asia Pacific Colorectal Screening score in Korea: An analysis of the Cancer Screening Cohort. *Epidemiol. Health* **2021**, *43*, e2021069. [[CrossRef](#)]
67. Liu, Y.; Colditz, G.A.; Rosner, B.A.; Dart, H.; Wei, E.; Waters, E.A. Comparison of Performance Between a Short Categorized Lifestyle Exposure-based Colon Cancer Risk Prediction Tool and a Model Using Continuous Measures. *Cancer Prev. Res.* **2018**, *11*, 841–848. [[CrossRef](#)] [[PubMed](#)]
68. 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. *Chin. J. Cancer Res.* **2014**, *26*, 4–11. [[CrossRef](#)] [[PubMed](#)]
69. Kaminski, M.F.; 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*, 1112–1119. [[CrossRef](#)] [[PubMed](#)]
70. Kim, D.H.; Cha, J.M.; Shin, H.P.; Joo, K.R.; Lee, J.I.; Park, D.I. Development and Validation of a Risk Stratification-Based Screening Model for Predicting Colorectal Advanced Neoplasia in Korea. 2015. Available online: www.jcge.com (accessed on 30 November 2022).
71. Cooper, J.A.; Ryan, R.; Parsons, N.; Stinton, C.; Marshall, T.; Taylor-Phillips, S. The use of electronic healthcare records for colorectal cancer screening referral decisions and risk prediction model development. *BMC Gastroenterol.* **2020**, *20*, 78. [[CrossRef](#)] [[PubMed](#)]
72. Jung, Y.S.; Park, C.H.; Kim, N.H.; Lee, M.Y.; Park, D.I. Impact of Age on the Risk of Advanced Colorectal Neoplasia in a Young Population: An Analysis Using the Predicted Probability Model. *Dig. Dis. Sci.* **2017**, *62*, 2518–2525. [[CrossRef](#)]
73. Auge, J.M.; Pellise, M.; Escudero, J.M.; Hernandez, C.; Andreu, M.; Grau, J.; Buron, A.; López-Cerón, M.; Bessa, X.; Serradesanferm, A.; et al. Risk Stratification for Advanced Colorectal Neoplasia According to Fecal Hemoglobin Concentration in a Colorectal Cancer Screening Program. *Gastroenterology* **2014**, *147*, 628–636.e1. [[CrossRef](#)] [[PubMed](#)]
74. Deng, J.; Zhou, Y.; Dai, W.; Chen, H.; Zhou, C.; Zhu, C.; Ma, X.; Pan, S.; Cui, Y.; Xu, J.; et al. Noninvasive predictive models based on lifestyle analysis and risk factors for early-onset colorectal cancer. *J. Gastroenterol. Hepatol.* **2023**, *38*, 1768–1777. [[CrossRef](#)]
75. Arnau-Collell, C.; Díez-Villanueva, A.; Bellosillo, B.; Augé, J.M.; Muñoz, J.; Guinó, E.; Moreira, L.; Serradesanferm, A.; Pozo, À.; Torà-Rocamora, I.; et al. Evaluating the Potential of Polygenic Risk Score to Improve Colorectal Cancer Screening. *Cancer Epidemiol. Biomark. Prev.* **2022**, *31*, 1305–1312. [[CrossRef](#)]
76. Klooster, C.C.V.; Ridker, P.M.; Cook, N.R.; Aerts, J.G.; Westerink, J.; Asselbergs, F.W.; van der Graaf, Y.; Visseren, F.L.; Nathoe, H.; de Borst, G.; et al. Prediction of Lifetime and 10-Year Risk of Cancer in Individual Patients With Established Cardiovascular Disease. *JACC CardioOncol.* **2020**, *2*, 400–410. [[CrossRef](#)]

77. Wei, E.K.; Colditz, G.A.; Giovannucci, E.L.; Fuchs, C.S.; 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. [[CrossRef](#)]
78. Wei, E.K.; Colditz, G.A.; Giovannucci, E.L.; Wu, K.; Glynn, R.J.; Fuchs, C.S.; Stampfer, M.; Willett, W.; Ogino, S.; Rosner, B. A Comprehensive Model of Colorectal Cancer by Risk Factor Status and Subsite Using Data From the Nurses' Health Study. *Am. J. Epidemiol.* **2017**, *185*, 224–237. [[CrossRef](#)]
79. Freedman, A.N.; Slattery, M.L.; Ballard-Barbash, R.; Willis, G.; Cann, B.J.; Pee, D.; Gail, M.H.; Pfeiffer, R.M. Colorectal Cancer Risk Prediction Tool for White Men and Women Without Known Susceptibility. *J. Clin. Oncol.* **2009**, *27*, 686–693. [[CrossRef](#)] [[PubMed](#)]
80. Huang, J.L.; Chen, P.; Yuan, X.; Wu, Y.; Wang, H.H.; Wong, M.C. An algorithm to predict advanced proximal colorectal neoplasia in Chinese asymptomatic population. *Sci. Rep.* **2017**, *7*, 46493. [[CrossRef](#)] [[PubMed](#)]
81. Riley, R.D.; Debray, T.P.A.; Collins, G.S.; Archer, L.; Ensor, J.; van Smeden, M.; Snell, K.I.E. Minimum sample size for external validation of a clinical prediction model with a binary outcome. *Stat. Med.* **2021**, *40*, 4230–4251. [[CrossRef](#)] [[PubMed](#)]
82. Song, M.; Emilsson, L.; Roelstraete, B.; Ludvigsson, J.F. Risk of colorectal cancer in first degree relatives of patients with colorectal polyps: Nationwide case-control study in Sweden. *BMJ* **2021**, *373*, n877. [[CrossRef](#)] [[PubMed](#)]
83. Ramspek, C.L.; Teece, L.; Snell, K.I.E.; Evans, M.; Riley, R.D.; van Smeden, M.; van Geloven, N.; van Diepen, M. Lessons learnt when accounting for competing events in the external validation of time-to-event prognostic models. *Leuk. Res.* **2022**, *51*, 615–625. [[CrossRef](#)] [[PubMed](#)]
84. Hingorani, A.D.; Windt, D.A.V.D.; Riley, R.D.; Abrams, K.; Moons, K.G.M.; Steyerberg, E.W.; Schroter, S.; Sauerbrei, W.; Altman, D.G.; Hemingway, H.; et al. Prognosis research strategy (PROGRESS) 4: Stratified medicine research. *BMJ* **2013**, *346*, e5793. [[CrossRef](#)]
85. Vickers, A.J.; van Calster, B.; Steyerberg, E.W. A simple, step-by-step guide to interpreting decision curve analysis. *Diagn. Progn. Res.* **2019**, *3*, 18. [[CrossRef](#)]
86. Vickers, A.J.; Elkin, E.B. Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. *Med. Decis. Mak.* **2006**, *26*, 565–574. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.