



Diagnostic Risk Prediction Models for Upper Gastrointestinal Cancers: A Systematic Review

Tyler Seyhan Saunders, Pawandeep Virpal, Maria Andreou, Asha Parmar, Christina Derksen, Oleg Blyuss, Fiona M. Walter, and Garth Funston

ABSTRACT

Upper gastrointestinal (UGI) cancers are often detected late. Risk prediction models could facilitate earlier detection by identifying patients at risk for further investigation. We systematically reviewed evidence on UGI diagnostic risk prediction models. A search of MEDLINE, Embase, and CENTRAL was conducted for studies reporting on the development and/or validation of diagnostic risk prediction models for UGI cancers (pancreatic, gastric, esophageal, gallbladder, and/or biliary tract). Studies had to report at least one quantitative measure of model performance to be eligible for inclusion. A total of 82 studies describing 162 UGI risk models were included. Models predicted gallbladder ($n = 6$), gastric ($n = 25$), esophageal ($n = 34$),

gastroesophageal ($n = 14$), and pancreatic ($n = 83$) cancers. Most models used logistic regression, but machine learning was increasingly used from 2019. In total, 366 unique variables were incorporated across models. Only 33 models were externally validated, with 15 achieving an AUC ≥ 0.80 . This review highlights that several models perform well in predicting UGI cancers on external validation. Future research is needed to compare the best-performing models and assess their clinical utility, acceptability, and cost-effectiveness. Given the significant overlap in at-risk populations and predictors across UGI cancers, there may also be scope to develop UGI “multicancer” models.

Introduction

Upper gastrointestinal (UGI) cancers, including pancreatic, gastric, esophageal, gallbladder, and biliary tract cancers, remain significant contributors to global cancer mortality, with poor survival outcomes worldwide (1). Pancreatic cancer, for example, has one of the lowest survival rates of any cancer, with a five-year net survival rate of less than 10% in England (2) and a five-year relative survival rate of 13% in the United States (3). Gastric, esophageal, gallbladder, and biliary tract cancers all have five-year survival rates of 36% or less in England (2, 4, 5) and the United States (6–8). Late-stage diagnosis is a major factor contributing to poor outcomes. For example, localized pancreatic cancer has a three-year net survival of over 25% in England, compared with 15% for regional and 1% for distant pancreatic cancer (9). However, most pancreatic cancers are diagnosed at a distant stage (9).

Some national guidelines advocate screening for pancreatic cancer in higher-risk groups (10, 11), whereas several countries with high incidence rates offer screening programs for esophageal and gastric cancer (12, 13). But, in most countries, screening programs for asymptomatic individuals at average risk of UGI cancers are not recommended. Most patients with UGI cancer are diagnosed after they develop symptoms and present to their doctor. However, most

symptoms of UGI cancers, such as weight loss, abdominal pain, and fatigue (14, 15), are nonspecific and easily misattributed to benign conditions by both patients and clinicians (16, 17). This poses a diagnostic challenge. Earlier diagnosis of UGI cancers, either through asymptomatic screening or earlier identification of cancer in patients with nonspecific symptoms, could improve outcomes by allowing for timely intervention, expanding treatment options, and potentially improving long-term survival and enhancing quality of life (18).

To address the challenges of late diagnosis, various risk prediction models have been developed to quantify an individual’s risk of UGI cancer and guide both surveillance and testing decisions. These models incorporate a range of established risk factors—including demographics, symptoms, novel biomarkers, routine lab tests, imaging, medications, and genetic data—and are intended to accurately quantify cancer risk and identify patients for further investigation. Increasingly, models are being integrated into the clinical decision-making processes. QCancer and electronic risk assessment tools (eRATs) are available to support decisions on further investigations in symptomatic individuals in UK clinical practice (19, 20), whereas new algorithms, such as the Enriching New-Onset Diabetes for Pancreatic Cancer (ENDPAC) model (21), are being used in clinical screening trials (22). Accurate prediction tools for UGI cancer could aid risk stratification, helping to select appropriate patients for further investigation based on individualized risk, avoiding potentially unnecessary tests, referrals, and harm while facilitating more timely investigation in patients with cancer. However, the utility of risk prediction models in identifying undiagnosed UGI cancers hinges on their accuracy and applicability in real-world settings. Additionally, given the overlap in risk factors, symptoms, and shared diagnostic pathways for the major UGI cancers, there is a rationale for examining these models collectively rather than in isolation as previous systematic reviews have (23–26). Recent evidence indicates that although the risk of individual cancers in those with symptoms or abnormal tests may be low, the combined risk of cancers may be above national thresholds for further investigation (27–30). Furthermore, multicancer early

Centre for Cancer Screening, Prevention and Early Diagnosis, Wolfson Institute of Population Health, Queen Mary University of London, London, United Kingdom.

Corresponding Author: Garth Funston, Wolfson Institute of Population Health, Charterhouse Square, London EC1M 6BQ, United Kingdom. E-mail: g.funston@qmul.ac.uk

Cancer Epidemiol Biomarkers Prev 2025;34:1240–51

doi: 10.1158/1055-9965.EPI-24-1714

This open access article is distributed under the Creative Commons Attribution 4.0 International (CC BY 4.0) license.

©2025 The Authors; Published by the American Association for Cancer Research

detection (MCED) tests (31) offer a promising approach in identifying multiple cancer types, particularly as reported performance in UGI cancers is high, which could enhance detection in patients presenting with nonspecific symptoms.

This review aims to identify existing diagnostic prediction models for each UGI cancer (and combinations of UGI cancers), to summarize how these models were developed and the predictors included across cancer types, and to critically examine model performance.

Materials and Methods

This review was prospectively registered on PROSPERO (CRD42024511177) and is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (32).

Search strategy

An electronic literature search of MEDLINE, Embase, and CENTRAL was performed from January 1, 2000, to October 2023 for primary studies. The search strategy combined keywords and Medical Subject Headings (MeSH) related to UGI cancers, risk prediction models, and statistical/machine learning (ML) methods (full search terms can be found in Supplementary Table S1). Reference lists of studies and reviews were manually searched to identify additional suitable studies. No language restrictions were applied. Only studies published in peer-reviewed journals were eligible for inclusion, and conference abstracts were excluded.

Eligibility criteria

Studies that fulfilled the following inclusion criteria were included: (i) the study described the development and/or validation of a multivariate diagnostic prediction model designed to identify patients with a UGI cancer, (ii) the study investigated any UGI cancer alone or in combination, (iii) the study reported at least one quantitative measure of model performance (discrimination, calibration, or threshold accuracy metrics), and (iv) the majority of the sample comprised individuals above 18 years old. For the purpose of this review, a UGI cancer was defined as one of the following: pancreatic, esophageal, gallbladder, gastric, or extrahepatic biliary tract cancer. Studies were excluded if (i) the study focused solely on prognostic models or models aiming to provide an indication of relapse or recurrence; (ii) the reported model did not include a patient-related variable such as demographics, symptoms, or risk factors (to exclude models that were focused exclusively on multi-panel biomarker tests); and (iii) the study was a nonprimary study, not peer-reviewed, or was a conference abstract. All study designs and recruitment settings were eligible for inclusion.

Independent researchers (T.S. Saunders, M. Andreou, A. Parmar) screened studies for inclusion/exclusion and resolved any discrepancies through discussion.

Data extraction

Data were extracted from eligible studies using Covidence software by pairs of independent researchers (T.S. Saunders, M. Andreou, A. Parmar, P. Virpal). This included study characteristics (year, country, design, setting), sample characteristics (recruitment, number of participants, sex, ethnicity, age, UGI cancer type), model characteristics (candidate predictors, predictors included in the final model, development method, missing data), and model performance measures (discrimination measure, any validation methods,

calibration metrics, performance by subgroups: age/sex/ethnicity/geographical location/socioeconomic status). The primary outcome was a UGI cancer of interest. Data were also extracted where a study reported on model performance differences between different population groups such as ethnicities, socioeconomic status, and sex, to examine whether models may exacerbate any existing health disparities between population groups. Data extraction was conducted using an extraction template, which was piloted before use. Where a study described multiple models, data for each were extracted separately. Any discrepancies were resolved through discussion.

TRIPOD classification

Studies were classified according to the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) criteria (33). The criteria were developed with the aim of improving the transparency of the reporting of risk prediction model studies and provide a useful framework for categorizing risk prediction model studies. Studies can be classified into the following categories: 1a (development only), 1b (development and validation using resampling), 2a (random split-sample development and validation), 2b (nonrandom split-sample development and validation), 3 (development and validation using separate data), and 4 (validation only).

Risk of bias

The Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2; ref. 34) tool was used to evaluate the risk of bias and applicability of included studies. The QUADAS-2 tool consists of four key domains: patient selection, index test, reference standard, and flow and timing. Each domain is assessed in terms of risk of bias, receiving a rating of “high,” “low,” or “unclear” (where insufficient information is provided to make a judgment). Signaling questions are included to assist in judgments about risk of bias. The first three domains also receive a rating of “high,” “low,” or “unclear” for concerns regarding applicability. The risk of bias was assessed independently by two reviewers using the QUADAS-2 tool. Any discrepancies were resolved by discussion.

Narrative synthesis

Model performance metrics were selected as the outcome measure of the review, with AUC chosen as the primary measure as previous reviews have identified it as the most commonly reported outcome (23, 24, 26). If AUC was not reported, sensitivity and specificity were chosen instead. Meta-analysis was not possible due to heterogeneity in terms of study design, study setting, model development methods, and sample characteristics. Therefore, studies were grouped by study setting and UGI cancer. This grouping was necessary as the characteristics of samples recruited from different settings may influence performance, with clinical samples more likely to report symptoms than population-based samples.

There was variation in the categorization of gastroesophageal cancers, with eight studies including it as a single category, whereas others treated esophageal cancer and gastric cancer as distinct entities. This inconsistency in categorization could impact the comparability of models and outcomes across studies, and so gastroesophageal was treated as a separate UGI category in the current review. Performance metrics were compared in tabular and graphical form and summarized with a narrative synthesis. Externally validated models (per TRIPOD criteria) were emphasized in the analysis. Models were considered unique if they included

different variables, employed different statistical methods, or aimed to detect different stages of UGI cancer. If an existing model was updated with new variables, this was classified as a unique model. Model development methods were grouped into either “ML” or “conventional statistical” methods. For the purpose of this review, any method more advanced than logistic regression was classified as “ML,” whereas logistic regression itself was classified as a conventional statistical method.

Data availability

Data sharing is not applicable to this article as no data were created in this systematic review.

Results

Search results

After the initial search, 7,941 studies were identified for screening. A further 43 studies were identified from manual reference list searching. After the removal of duplicates, titles and abstracts of 6,010 studies were screened. After full-text screening of 187 studies, 82 studies met the inclusion criteria (see Fig. 1).

Study characteristics

Eighty-two studies included in the current review (see Supplementary Table S2) focused on gallbladder ($n = 6$), gastric ($n = 18$), esophageal ($n = 14$), gastroesophageal ($n = 8$), and pancreatic ($n = 33$) cancers—with three studies focusing on both pancreatic and gastroesophageal cancers. Most studies were conducted in China ($n = 35$), followed by the United Kingdom ($n = 11$) and the United States ($n = 9$).

A total of 162 unique models were described: six for gallbladder, 25 for gastric, 34 for esophageal, 14 for gastroesophageal, and 83 for pancreatic cancer. Only one model targeted multiple UGI cancers simultaneously (esophageal, pancreatic, and gastric cancers; ref. 35), and none focused on extrahepatic biliary cancer.

According to TRIPOD criteria, most studies were classified as 2a (random split-sample development and validation, $n = 24$, 29.3%), followed by 1b (development and validation using resampling, $n = 20$, 23.2%) and 1a (development only, $n = 18$, 22%). Other classifications included 3 (development and validation using separate data, $n = 14$, 17.1%), 4 (validation only, $n = 5$, 7.3%), and 2b (nonrandom split-sample development and validation, $n = 1$, 1.2%).

Study designs predominantly included case-control ($n = 49$) and cohort ($n = 33$) designs. Most were conducted in hospital or clinic settings ($n = 61$), with others in population-level settings ($n = 13$) and primary care ($n = 5$). Data sources varied, with most studies using existing health records ($n = 39$), followed by in-person interviews ($n = 11$) and questionnaires ($n = 9$).

Risk prediction model development

Model development methods varied across studies. Of all included models, 123 were developed using conventional statistical methods (Fig. 2A), most commonly logistic regression ($n = 104$), followed by Cox regression ($n = 8$) and logistic generalized linear models ($n = 7$; Fig. 2B). Thirty-nine models were developed using ML, such as ensemble learning techniques ($n = 20$) and neural networks ($n = 15$; Fig. 2C). Consistent with expectations, the use of ML increased over time, with the first ML models appearing in 2019, compared with conventional statistical models, which have been published since at least 2005.

Missing data

Of all included studies, only 17 studies reported the proportion of missing data for variables of interest (21, 36–51) although 31 studies reported how missing data were handled. Complete-case analysis was used in 17 studies (39–42, 44, 46, 48, 50, 52–58), in which patients were excluded if they had missing data for variables of interest. The remaining studies reported that missing values were imputed using multiple imputation (19, 20, 35, 37, 38, 43, 45, 47, 49, 59–62).

Risk prediction model variables

A total of 366 unique variables were included across models, grouped into 12 categories: biomarkers (including novel biomarkers), comorbidities, demographics, family history, genetics (e.g., SNPs, polygenic risk scores), general health (e.g., perceived health status), imaging, standard lab tests (e.g., platelet counts), lifestyle (e.g., smoking, alcohol, and diet), medications, symptoms, and other variables (e.g., cytological variables). Supplementary Table S3 lists all unique variables in each category, while Supplementary Table S4 provides details on variable selection and cutoffs/categorizations of variables in the 15 best-performing externally validated models.

Figure 3A–F illustrate the proportion of variable categories by UGI cancer type. Several variables were shared across models for different UGI cancers. Notably, three variables—age, sex, and alcohol consumption—are included in at least one model for each UGI cancer. Ethnicity, smoking, and body mass index (BMI) are present in models for all UGI cancers except gallbladder cancer. Additionally, CA125 and carcinoembryonic antigen (CEA) are included in at least one model for all UGI cancers except for gastroesophageal models. Conversely, some variable categories differed by cancer type; for example, imaging variables were more common in gallbladder models, whereas symptom variables were prevalent in gastroesophageal models but absent from gallbladder and gastric models.

Validation

Of the 162 risk prediction models identified, 38 were described without any validation, 91 models had been internally validated, and 33 had been externally validated.

A random split-sample validation was the most commonly employed internal validation technique, in which a model was developed on a subset of the total sample and tested on the remaining sample ($n_{models} = 44$; refs. 19, 20, 46–48, 51, 55, 57, 59, 60, 62–70). Cross-validation was the second most common internal validation method ($n_{models} = 36$; refs. 40, 41, 43–45, 49, 61, 71–81), followed by bootstrapping ($n_{models} = 17$; refs. 36, 53, 82–88).

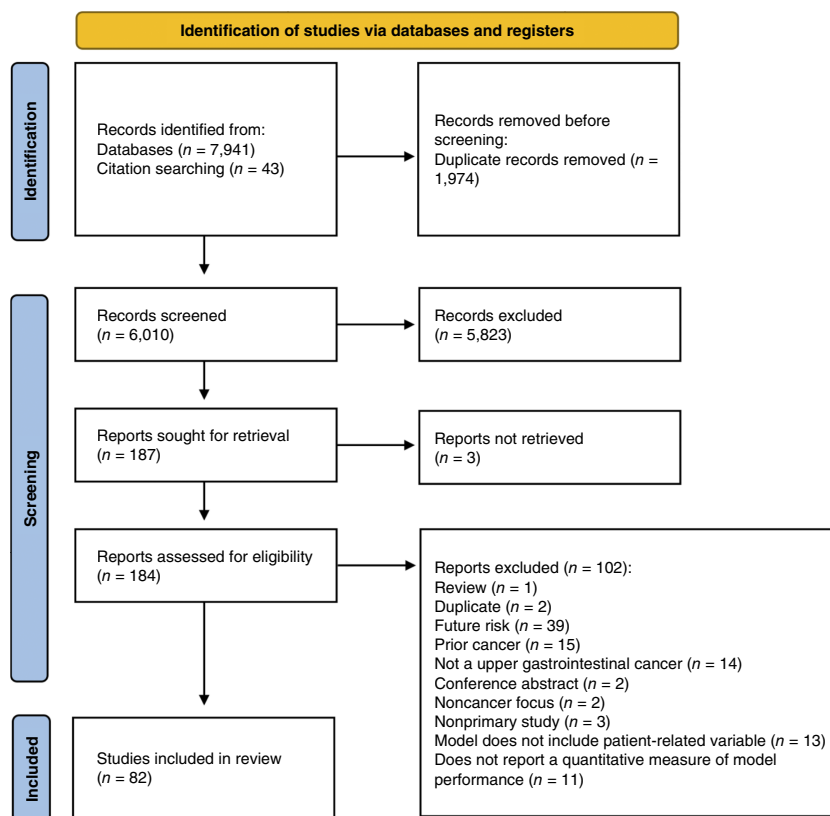
Nineteen studies reported the external validation of 33 different models (see Table 1). Thirteen studies reported the development of a prediction model and its evaluation on separate data (35, 37, 58, 61, 73, 88–95), whereas six studies evaluated the performance of an existing prediction model on separate data (50, 52, 54, 96–98). Of these studies, five were conducted at the population level, and 14 were conducted in a hospital/clinic setting.

Risk prediction model performance

Model discrimination, as measured by AUC, was reported for 152 models, with 134 (88.2%) reporting moderate-to-strong performance. Of the 33 externally validated models, 30 reported AUCs, and 15 demonstrated moderate-to-strong performance.

Figure 1.

Flow diagram of search strategy. Diagram showing the number of titles/abstracts screened, full texts reviewed, and studies included.



Performance across statistical categories

Among the externally validated models, 26 were developed with conventional statistics, and seven were developed using ML. ML models reported a slightly higher median AUC (0.82, IQR = 0.13) than conventional methods (AUC = 0.77, IQR = 0.11) although direct comparisons were not possible.

Performance of externally validated models in hospital-based settings

Twenty-two models were externally evaluated in hospital-based settings (35, 37, 51, 52, 58, 61, 88, 91–97). Below is a summary of models with the highest AUCs for each UGI cancer type (Fig. 4).

Gallbladder cancer: Two models achieved high performance. Zhou and colleagues (88) developed a model in patients with chronic cholecystitis, incorporating demographics, lifestyle, and imaging variables, achieving an AUC of 0.86. Another model included demographics, gallstone history, and blood tests (CEA, CA19-9), with an AUC of 0.88 [95% confidence interval (CI), 0.82–0.94; ref. 58].

Gastric cancer: Among 11 externally validated models (35, 50, 91, 94, 95), AUCs ranged from 0.60 to 0.88. Tao and colleagues (94) achieved the highest AUC (0.88) with a model, including demographics, history of *Helicobacter pylori*, lab tests (serum pepsinogen), and lifestyle factors (type of drinking water) in a very small cohort ($n = 26$). Another notable model by Huang and colleagues included demographics, anemia status, comorbidities, and smoking, achieving an AUC of 0.82 (91).

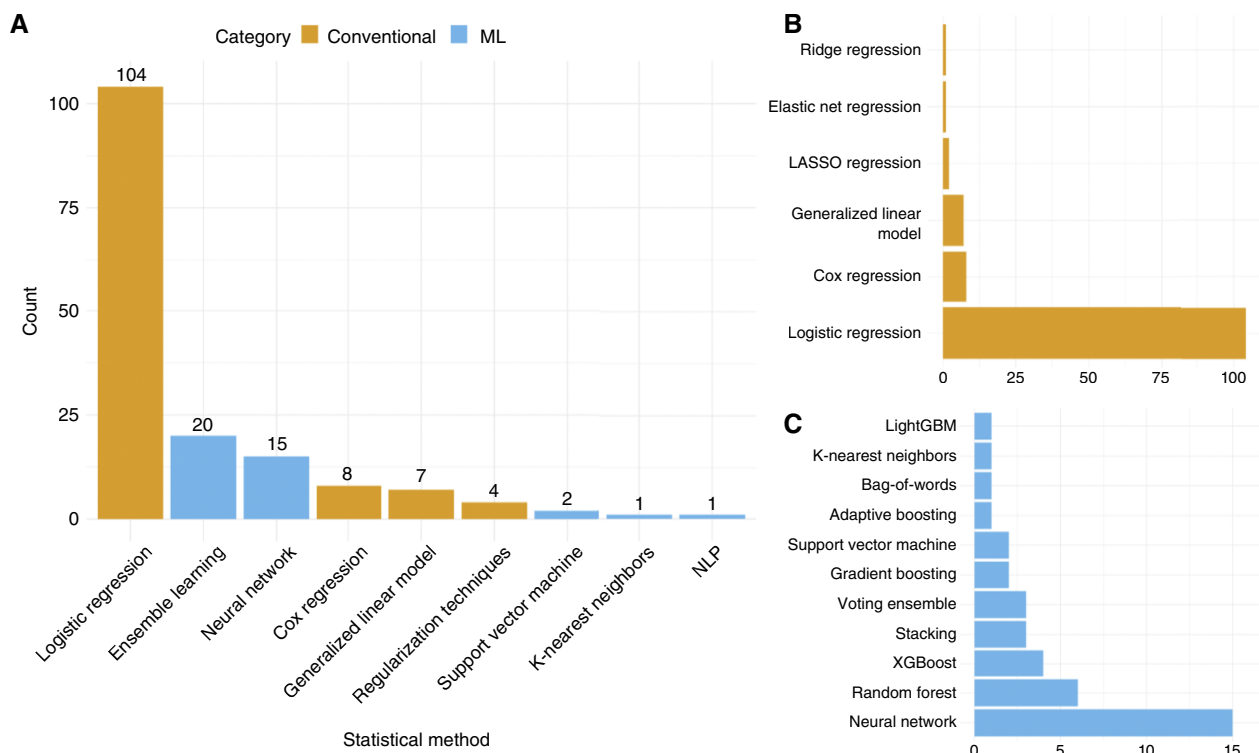
Gastroesophageal cancer: Only one model was externally validated, developed by Kapoor and colleagues (92) for patients referred to a rapid access UGI cancer service. The model, based on symptoms (e.g., dysphagia, weight loss) and age, achieved a high sensitivity (92.3%) but poor specificity (32.7%) at the selected threshold.

Esophageal cancer: Four externally validated models (35, 37, 61, 96) reported AUCs ranging from 0.81 to 0.9. The highest AUC (0.96) was reported by Gao and colleagues for a model developed in patients undergoing UGI endoscopy screening (37). It included demographics, lifestyle factors (e.g., smoking, pickled food), family history, and cytological features. Ho and colleagues (61) also reported a high AUC of 0.92 with a model including demographics, lifestyle factors (e.g., smoking, diet), comorbidities, and symptoms.

Pancreatic cancer: Four externally validated models reported AUCs ranging from 0.77 to 0.91 (35, 52, 93). Luan and colleagues (35) developed a model incorporating blood biomarkers (e.g., α -fetoprotein, CA19-9, and CEA) and demographics, achieving an AUC of 0.91. Henriksen and colleagues (52) updated an existing gene-based model (71) to include serum CA19-9, resulting in an AUC of 0.85 (95% CI, 0.79–0.91), although calibration was reported to be poor.

Performance of externally validated models in population-based settings

Eleven models were externally validated in population-based settings across five studies (Fig. 4; refs. 54, 73, 89, 90, 98).

**Figure 2.**

Statistical methods used to develop each model. **A**, Bar chart showing the number of models developed with different methods, either ML (blue) or conventional statistics (orange). **B**, The most common methods used for conventional statistics and the number of models. **C**, The most common methods used for ML techniques and the number of models.

Gastric cancer: Hu and colleagues (98) validated a model for high-risk individuals in China, incorporating lab tests (pepsinogen I/II ratio, gastrin-17), lifestyle variables (diet), and demographics, achieving an AUC of 0.71.

Esophageal cancer: Li and colleagues described a model in high-risk areas of China, incorporating demographics, family history, lifestyle (e.g., smoking, diet), symptoms, and comorbidities, with an AUC of 0.80 (95% CI, 0.75–0.84; ref. 54). Another model included similar variables such as demographics, family history, and diet, with an AUC of 0.71 (95% CI, 0.65–0.78; ref. 73).

Pancreatic cancer: Eight models were externally validated across two studies (89, 90), with AUCs ranging from 0.69 to 0.83. Placido and colleagues used a deep learning algorithm on electronic health record data and developed a model using demographics and comorbidities (specifically, International Classification of Diseases diagnostic codes) to achieve an AUC of 0.83 when predicting 6 months prior to diagnosis (95% CI, 0.83–0.83) and 0.71 (95% CI, 0.71–0.71) 36 months prior to diagnosis. Salvatore and colleagues described a model using UK Biobank data, incorporating demographics, comorbidities (International Classification of Diseases codes), lifestyle (e.g., smoking, alcohol), and a polygenic risk score for pancreatic cancer, reaching an AUC of 0.81 (95% CI, 0.79–0.82).

Performance differences across different population groups

Only two studies examined differences by population subgroups although neither was an external validation study. Chen and

colleagues (65) reported an internally validated pancreatic cancer model that performed better in females (AUC = 0.85) than in males (AUC = 0.82) and showed slightly improved performance for Black (AUC = 0.84) over White participants (AUC = 0.83). Similarly, Rubenstein and colleagues (47) reported an internally validated gastroesophageal cancer model, which performed better in females (AUC = 0.83) than in males (AUC = 0.72). This model also performed better in Asian (AUC = 0.80) than in White (AUC = 0.75) and Black (AUC = 0.75) participants.

Calibration

Calibration metrics were reported for 22 of the 33 externally validated models (37, 50, 52, 54, 58, 73, 88, 90, 91, 93, 94). The most common method of assessing calibration was the Hosmer–Lemeshow test ($n = 14$), followed by calibration plots ($n = 8$) and the Brier score ($n = 1$). In studies reporting calibration metrics, all but two models were reported to be well-calibrated. Henriksen and colleagues (52) reported poor calibration for both included models.

Risk of bias

The main sources of bias were identified in the “patient selection” and “index test” domains (Supplementary Fig. S1). Fifty-three studies were flagged as being at high risk of bias for the patient selection domain due to the use of a case–control design. Furthermore, within the index test domain, 72 studies were flagged as being potentially at high risk of bias due to failing to predefine the tool

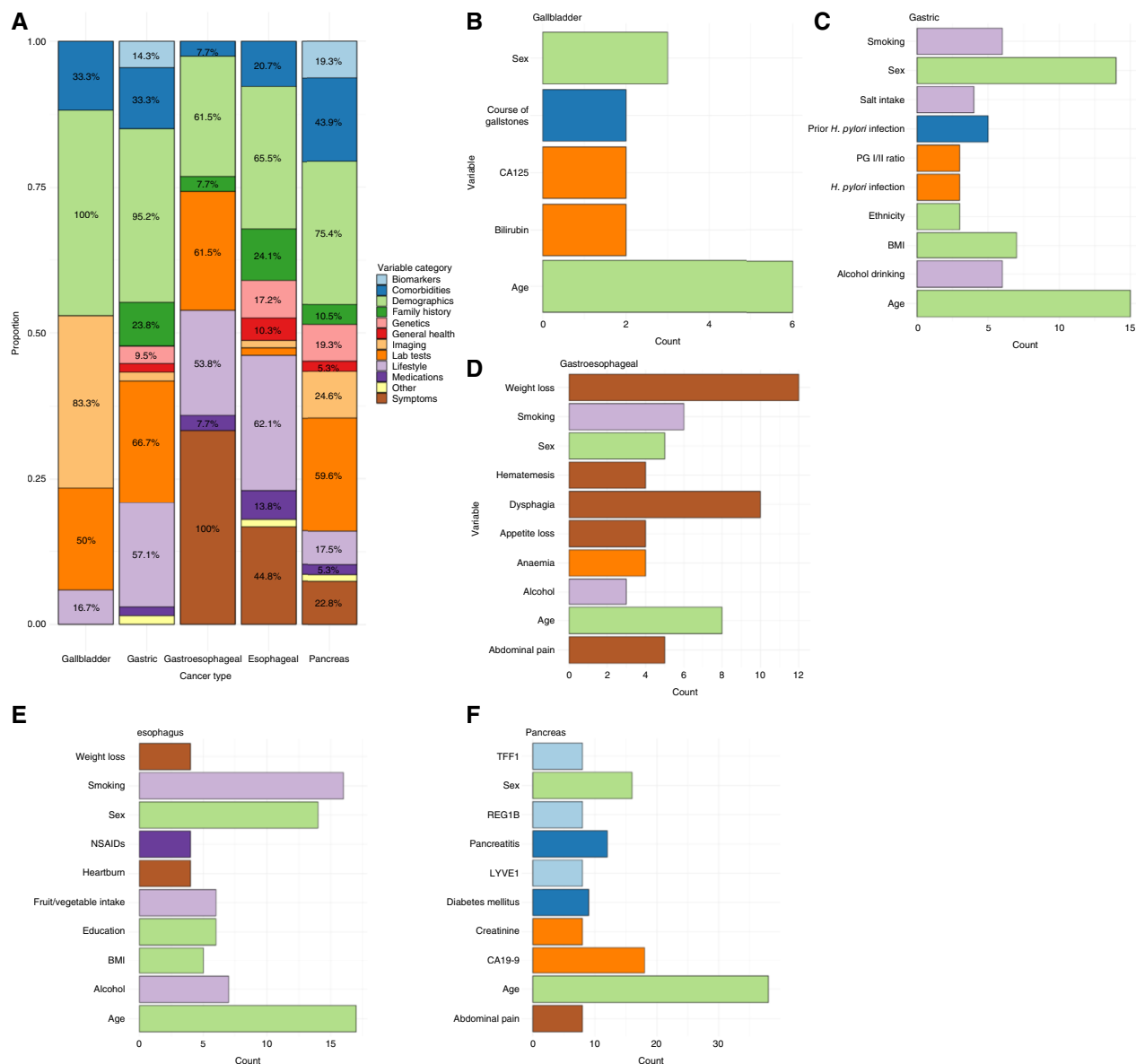


Figure 3. Variables included in models across UGI cancer types. **A**, Stacked bar chart showing the proportion of models containing variables of each category by UGI cancer type. **B–F**, Up to the top 10 most common variables included in models for each cancer type: **(B)** gallbladder, **(C)** gastric, **(D)** gastroesophageal, **(E)** esophageal, and **(F)** pancreatic. Data shown are the number of models containing each variable.

threshold. Overall, the risk of bias for the “reference standard” and “flow and timing” domains was judged to be low. However, 20 studies were judged to be potentially at high risk of bias in the “flow and timing” domain as not all patients were included in some analyses. The applicability of the risk of bias was judged to be low for most studies.

Discussion

We conducted a systematic review of diagnostic risk prediction models for UGI cancers, including gallbladder, gastric, esophageal, and pancreatic cancers. This review identified 82 studies describing

162 risk prediction models, several of which had been externally validated and demonstrated good performance. To our knowledge, this is the first comprehensive systematic review of risk prediction models across UGI cancers.

Most models in this review were developed using traditional statistical methods, primarily logistic regression. However, there has been an increase in ML approaches such as neural networks and random forest since 2019—reflecting the broader adoption of ML in risk prediction research. ML offers potential advantages, including improved predictive accuracy through capturing complex, nonlinear relationships among predictors. However, these methods also pose challenges, requiring larger datasets, advanced computational

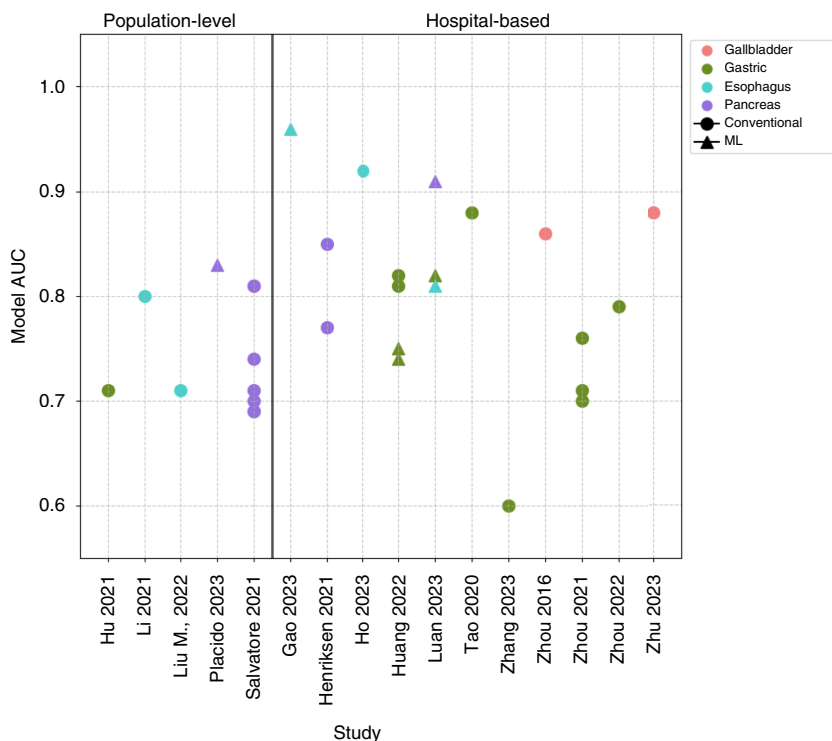


Figure 4.

AUCs reported in external validation studies. Dot plot showing the AUC reported by each external validation study, separated by study setting. Each data point represents a single model. Data points are colored according to UGI cancer site (red, gallbladder; green, gastric; blue, esophageal; purple, pancreatic). Data point shaped corresponds to the model development method (circle, conventional statistics; triangle, ML).

resources, and specialized expertise in model tuning and validation. In this review, most externally validated models showed moderate-to-strong performance, regardless of methodological approach. Few studies compared the performance of models developed with conventional versus artificial intelligence (AI) methods (37, 49, 55, 62, 63, 66, 75, 91). Although two found their ML approach outperformed conventional approaches (55, 62), the rest reported comparable performance. Although few studies, and only two external validation studies, directly compared ML and conventional statistical models, such comparisons could provide valuable insights. Future research should focus on comparing model performance across methods and evaluating their interpretability and clinical utility.

Most models in this review were developed to detect a single UGI cancer, likely reflecting the structure of clinical guidelines, screening programs, and diagnosis pathways, which are generally cancer-site specific (99). This single-cancer approach in risk prediction model development could limit the potential for multicancer detection (MCD) approaches. To address the limitations of site-specific pathways, several countries, including Denmark and the UK, have introduced nonspecific symptom pathways, which adopt a multicancer approach for patients presenting with symptoms, which do not clearly indicate a single cancer type. The development of blood-based MCD tests, such as those used in the NHS-Galleri trial (100), further highlights this shift, especially for UGI cancers in which overlapping risk factors and nonspecific symptoms can complicate diagnosis.

Across the reviewed models, 366 unique variables were identified. Age, sex, and alcohol consumption were frequently included across models for each UGI cancer type. Ethnicity, smoking, and BMI were included in at least one model for all UGI cancers except gallbladder. These findings align with previous studies noting the prominence of age, sex, BMI, alcohol, and smoking in pancreatic and esophageal cancer models (23, 24). Laboratory tests were commonly

included in models across all UGI cancers except gastroesophageal cancer. However, notable differences in variable inclusion were also evident across cancer types. For example, imaging variables were frequently included in gallbladder cancer models but were rarely included in models for other UGI cancers, whereas symptoms were consistently included in gastroesophageal models but absent from gallbladder and gastric models.

Several variables were commonly included across high-performing externally validated models. For example, hospital-based gastric cancer models with the highest AUCs included sex and prior *H. pylori* infection (91, 94)—a finding consistent with another review (25). Similarly, high-performing hospital-based esophageal cancer models included age, sex, and smoking (37, 61), whereas the top-performing pancreatic models in population settings incorporated comorbidities such as pancreatitis and neoplasms of the digestive system (89, 90).

The overlap of common variables across UGI cancers underscores the potential for MCD strategies. Shared variables such as demographics (age, sex, BMI, ethnicity), lifestyle (alcohol, smoking), and symptoms could be used in MCD models to identify patients at high risk of several UGI cancers. Given that diagnostic pathways for UGI cancers frequently overlap, MCD models could streamline current processes and improve resource efficiency. They could also help increase the pretest probability for patients undergoing MCD tests. However, MCD models should balance shared variables with cancer-specific predictors (such as *H. pylori* infection for gastric cancer) to ensure robust performance across various UGI cancers.

Consistent with previous research (23–25), we found that a minority of prediction models (20%) had been externally validated, and few studies reported calibration metrics. It is well documented that model performance often declines when models are applied to

Table 1. Table of external validation studies.

Lead author (year)	Outcome	Model development	Variable categories included in the model										AUC (95% CI)
			D	C	Sx	Img	Lab	LS	Gen	GH	FH	Oth	
Population-based													
Hu (2021)	GC	LR	●				●	●					0.71 (0.64–0.77)
Li (2021)	OSCC	LR	●	●	●			●			●		0.80 (0.75–0.84)
Liu, M (2022)	OSCC	LR	●					●			●		0.71 (0.65–0.78)
Placido (2023)	PC	Neural network		●	●		●						0.83 (0.83–0.83) ^a
Salvatore (2021)	PC	GLM	●						●	●			0.69 (0.67–0.71)
Salvatore (2021)	PC	GLM	●						●	●			0.69 (0.69–0.71)
Salvatore (2021)	PC	GLM	●						●	●			0.71 (0.69–0.73)
Salvatore (2021)	PC	GLM	●	●					●	●			0.70 (0.69–0.73)
Salvatore (2021)	PC	GLM	●	●					●	●			0.81 (0.79–0.82)
Salvatore (2021)	PC	GLM	●	●					●	●			0.74 (0.72–0.77)
Salvatore (2021)	PC	GLM	●	●					●	●			0.81 (0.79–0.83)
Hospital-based													
Gao (2023)	OSCC (78%) OAC (22%)	Light-gradient boosting	●	●					●		●	●	0.96 (0.92–0.99)
Henriksen (2021)	PDAC	LR	●						●				0.77 (0.69–0.84)
Henriksen (2021)	PDAC	LR	●				●		●				0.85 (0.79–0.91)
Ho (2023)	OSCC (7%) OAC (93%)	LR	●	●	●				●				0.92 (0.88–0.96)
Huang (2022)	GC (non-cardia)	Random forest	●	●				●					0.74 (NR)
Huang (2022)	GC (non-cardia)	K-nearest neighbors	●	●				●					0.75 (NR)
Huang (2022)	GC (non-cardia)	LASSO regression	●	●				●	●				0.82 (NR)
Huang (2022)	GC (non-cardia)	LR	●	●				●	●				0.81 (NR)
Kapoor (2005)	G-O	LR			●			●					NR
Luan (2023)	GC	Random forest	●					●					0.82 (NR)
Luan (2023)	PC	Random forest	●					●					0.91 (NR)
Luan (2023)	OC	Random forest	●					●					0.81 (NR)
Shen (2013)	MPCN	LR					●	●					NR
Tao (2020)	GC	LR	●	●				●	●				0.88 (0.85–0.92)
Yokoyama (2013)	OSCC (95%) OAC (5%)	LR							●				NR
Zhang (2023)	GC	LR	●					●	●		●		0.60 (NR)
Zhou (2016)	GBC	LR	●	●		●			●				0.86 (NR)
Zhou (2021)	GC	LR	●						●		●		0.76 (0.71–0.82)
Zhou (2021)	GC	LR	●						●		●		0.71 (0.61–0.81)
Zhou (2021)	GC	LR	●						●		●		0.70 (0.57–0.83)
Zhou (2022)	GC	LR	●					●					0.79 (NR)
Zhu (2023)	GBC	LR		●		●		●					0.88 (0.82–0.94)

Abbreviations: C, comorbidities; D, demographics; DAC, pancreatic ductal adenocarcinoma; FH, family history; GBC, gallbladder cancer; GC, gastric cancer; Gen, genetics; GH, general health; GLM, generalized linear model; G-O, gastro-oesophageal cancer; Img, imaging; Lab, lab test; LASSO, least absolute shrinkage and selection operator; LR, logistic regression; LS, lifestyle; MPCN, malignant pancreatic cystic neoplasm; NR, not reported; OAC, oesophageal adenocarcinoma; OC, oesophageal cancer unspecified; OSCC, oesophageal squamous cell carcinoma; Oth, other; PC, pancreatic cancer unspecified; Sx, symptoms.

^aAUC = 0.83 (6-month prediction interval), 0.79 (12-month prediction interval), and 0.71 (36-month prediction interval).

external datasets. In this review, we found that while 82.7% of the models reviewed reported AUC values greater than 0.80, less than half of externally validated models reported AUCs ≥ 0.80 . This highlights the importance of external validation in ensuring that risk models maintain robust performance prior to clinical implementation.

Nevertheless, some externally validated models demonstrated strong performance. For example, the esophageal cancer model of Gao and colleagues (37) reported an AUC of 0.96, whereas Luan and colleagues (35) described a multicancer prediction model with an AUC of 0.91 for pancreatic cancer and 0.82 for gastric cancer. These examples demonstrate that certain models can retain high performance in external validation, highlighting the potential of some models for real-world implementation.

Models identified in this review were developed primarily to support UGI risk assessment in two contexts: asymptomatic screening and triage of symptomatic patients. Although evidence in most countries does not support screening for UGI cancers in average-risk individuals, there is evidence that screening for pancreatic cancer in higher-risk groups and for gastric cancer in higher-incidence populations may improve survival or reduce mortality (12, 101). Models could be used to identify individuals at higher risk of UGI cancers for targeted screening tests. This approach is exemplified by the externally validated ML model of Gao and colleagues (37), designed to select patients at higher risk of esophageal cancer for screening via endoscopy, and by the ENDPAC model developed by Sharma and colleagues (21), which is currently being used in clinical trials to select higher-risk individuals with new-

onset diabetes for pancreatic screening with CT (22). The deep learning algorithm of Placido and colleagues (89) offers an alternative approach, which is intended to identify higher-risk individuals for screening based on complex healthcare patterns (disease trajectories) within longitudinal healthcare records. By contrast, the multicancer OncoSeek algorithm of Luan and colleagues (89), the only externally validated multicancer prediction model identified in this review, was developed to identify asymptomatic patients at elevated risk of multiple cancers (including pancreatic and esophageal), providing a potential alternative to single-site screening strategies. In patients presenting to health care with nonspecific symptoms, models identified in this study could be used to estimate the risk of one or more UGI cancers to determine whether further cancer investigation is warranted or to triage patients for urgent investigation. This is exemplified by Ho and colleagues' (61) model designed to triage patients with symptoms of possible esophageal cancer for endoscopy and the pancreatic and gastroesophageal models of Hippisley-Cox and colleagues (59, 60) (available within consolidated QCancer clinical tools in the United Kingdom), designed to identify symptomatic patients in primary care for further investigation in line with national guidelines.

This systematic review is, to the best of our knowledge, the first to examine risk prediction models across multiple UGI cancers, rather than focusing on individual cancers. However, it is not without limitations. We employed a broad search strategy to capture a wide range of studies. Although this approach enabled a comprehensive assessment of the literature, the resulting heterogeneity in the included studies meant it was not possible to perform a meta-analysis. Additionally, although we identified overlapping variables between models, inclusion frequency does not necessarily equate to clinical importance or predictive power as model variables are often selected based on the study aims and research interests of the authors. Different subtypes of UGI cancer, such as pancreatic adenocarcinoma and pancreatic neuroendocrine tumors, may have different risk factors and may present in different ways, and their detection has different clinical implications for patients. However, a minority of validated studies provided detailed information on the pathological breakdown of included tumors or pathology-specific performance of models, which hampers the comparison of model performance and clinical implementation.

This review has several clinical implications for developing, validating, and implementing risk prediction models for UGI cancers. First, the limited external validation and inconsistent reporting of calibration highlight a critical gap in the literature. For risk prediction models to be clinically useful, they must undergo rigorous external validation in populations that reflect real-world clinical settings, including primary care, to ensure they are both accurate and generalizable to diverse patient populations (102). Furthermore, the lack of comparative studies on externally validated models for the same UGI cancers complicates the selection of clinically viable models. Future validation studies should aim to compare the performance of new models against that of established, high-performing models.

Integration into clinical workflows is another essential consideration. If models are not easily integrated, they may face adoption barriers, regardless of their accuracy (103). Previous studies and reviews have highlighted significant barriers to the implementation and clinical adoption of prediction models (105–111), including those for UGI cancers (104). The implementation of cancer tools and systems is hindered by a range of barriers at the point of care and within wider healthcare systems. These include regularity hurdles (as clinical prediction models require regulatory approval in

many jurisdictions; refs. 107–109), software issues (including integrating models within clinical information technology (IT) systems; (106–108), challenges around integration into clinical workflow, and acceptability and trust issues about model predictions among clinicians (105, 110). Widely used tools, such as QCancer tools in the United Kingdom, are generally integrated into IT systems and aligned with national referral guidelines. Despite this, clinicians have identified barriers to QCancer use, including the time required to use the tools (104, 105).

Although the performance of several AI-based models identified in this review is promising, the implementation of such models presents additional challenges, and we are not aware of any AI-based UGI prediction models currently approved for use in clinical practice. Such models may be computationally intensive, and their application in many healthcare settings would require substantial investment and development in secure IT systems (111). Concerns also exist about AI explainability and how to include information from AI prediction models into shared decision-making (112). Further consideration must also be given to how models, such as the promising deep learning model of Placido and colleagues (89), which utilizes longitudinal healthcare records, could be integrated with national screening programs, which are often run at a regional or national level. The focus of this review was not specifically on clinical implementation, but we noted that although some included studies acknowledged challenges to implementation, there was limited discussion on how these challenges could be addressed. Concerningly, some models require data that may not be routinely available in clinical settings, which could limit utility. A systematic review is currently underway to provide a framework for prediction model development and implementation research (113), including recommendations for developers, healthcare teams, and policymakers.

A key concern is the potential for risk prediction models to exacerbate existing health inequalities. Data used to develop models can often be skewed by existing structural and social inequalities, and applying these models to practice could exacerbate existing health inequalities (114). Only two studies examined performance differences between population groups (e.g., by sex and ethnicity). Most studies in this review were conducted predominantly in White or Asian populations (specifically within China). Inclusive model design and evaluation are essential in preventing the exacerbation of ethnic, socioeconomic, rural–urban, and other existing disparities in cancer detection and outcomes.

Conclusion

In conclusion, this review identified a large number of UGI cancer risk prediction models, most of which predicted a single UGI cancer. Although some variables overlapped across different UGI cancers, others were cancer-specific—suggesting that multicancer early detection models would benefit from including both shared and unique risk factors. ML models have become more common in recent years although their performance appears generally comparable to conventional statistical models—underscoring the need for head-to-head comparative studies to evaluate performance differences. A significant limitation observed in the field is the lack of external validation. Nonetheless, several externally validated models demonstrated promising performance and merit further investigation for clinical application. Future research should ensure that UGI cancer prediction models are developed using data from the population in which they are intended to be used and that they are externally validated in diverse populations and their performance compared head-to-head (“benchmarked”) against the best-

performing existing models. It is critical that developers consider user acceptability and how their models can support clinical workflow at an early stage. Given the complexity of new AI-based models, early consideration of how models can be integrated within healthcare IT systems is also imperative to ensure that translation into clinical practice is achievable.

Authors' Disclosures

G. Funston reports grants from Cancer Research UK and NIHR during the conduct of the study. No disclosures were reported by the other authors.

Acknowledgments

This study is funded by a Cancer Research UK programme grant "CANDETECT: Accelerating detection of upper gastrointestinal cancers using a multicancer

early detection platform in primary care" (EDDPGM-May22\100002; G. Funston, O. Blyuss, T.S. Saunders, C. Derksen, F.M. Walter) and an NIHR Invention for Innovation grant "Artificial Intelligence to support cancer early diagnosis in general practice" (NIHR207533; G. Funston). O. Blyuss acknowledges support from Barts Charity (G-001522). The views expressed are those of the authors and not necessarily those of Cancer Research UK, the NIHR, or the Department of Health and Social Care. The authors are grateful to Ben McGuirk for support in figure preparation.

Note

Supplementary data for this article are available at *Cancer Epidemiology, Biomarkers & Prevention Online* (<http://cebp.aacrjournals.org/>).

Received December 6, 2024; revised March 18, 2025; accepted May 20, 2025; posted first May 22, 2025.

References

- Allemani C, Matsuda T, Carlo V, Harewood R, Matz M, Nikšić M. Global surveillance of trends in cancer survival: analysis of individual records for 37,513,025 patients diagnosed with one of 18 cancers during 2000–2014 from 322 population-based registries in 71 countries (CONCORD-3). *Lancet* 2018; 391:1023–75.
- Office for National Statistics (ONS). Cancer survival in England: adult, stage at diagnosis and childhood - patients followed up to 2018. [cited 2025 Mar 4]. Available from: <https://www.ons.gov.uk/releases/cancersurvivalinenglandadultstageatdiagnosisandchildhoodpatientsfollowedupto2018>.
- American Cancer Society. Survival rates for pancreatic cancer. [cited 2025 Mar 4]. Available from: <https://www.cancer.org/cancer/types/pancreatic-cancer/detection-diagnosis-staging/survival-rates.html>
- Cancer Research UK. Survival of gallbladder cancer; 2023. [cited 2025 Mar 4]. Available from: <https://www.cancerresearchuk.org/about-cancer/gallbladder-cancer/survival>.
- Cancer Research UK. Survival for bile duct cancer; 2024. [cited 2025 Mar 4]. Available from: <https://www.cancerresearchuk.org/about-cancer/bile-duct-cancer/survival>.
- American Cancer Society. Bile duct cancer survival rates | cholangiocarcinoma survival rates. [cited 2025 Mar 4]. Available from: <https://www.cancer.org/cancer/types/bile-duct-cancer/detection-diagnosis-staging/survival-by-stage.html>.
- American Cancer Society. Stomach (gastric) cancer survival rates. [cited 2025 Mar 4]. Available from: <https://www.cancer.org/cancer/types/stomach-cancer/detection-diagnosis-staging/survival-rates.html>.
- American Cancer Society. Survival rates for esophageal cancer | esophageal cancer outlook. [cited 2025 Mar 4]. Available from: <https://www.cancer.org/cancer/types/esophagus-cancer/detection-diagnosis-staging/survival-rates.html>.
- Cabasag CJ, Arnold M, Rutherford M, Bardot A, Ferlay J, Morgan E, et al. Pancreatic cancer survival by stage and age in seven high-income countries (ICBP SURVMARK-2): a population-based study. *Br J Cancer* 2022;126:1774–82.
- NCCN. Genetic/familial high-risk assessment: breast, ovarian, pancreatic, and prostate. [cited 2025 Mar 4]. Available from: <https://www.nccn.org/guidelines/guidelines-detail?category=2&id=1545>.
- Aslanian HR, Lee JH, Canto MI. AGA clinical practice update on pancreas cancer screening in high-risk individuals: expert review. *Gastroenterology* 2020;159:358–62.
- Kim J-H, Han K-D, Lee JK, Kim H-S, Cha JM, Park S, et al. Association between the National Cancer Screening Programme (NSCP) for gastric cancer and oesophageal cancer mortality. *Br J Cancer* 2020;123:480–6.
- Sun D, Müller DT, Li Y, Nieboer D, Park JY, Suh M, et al. The effect of nationwide organized cancer screening programs on gastric cancer mortality: a synthetic control study. *Gastroenterology* 2024;166:503–14.
- Koo MM, Swann R, McPhail S, Abel GA, Elliss-Brookes L, Rubin GP, et al. Presenting symptoms of cancer and stage at diagnosis: evidence from a cross-sectional, population-based study. *Lancet Oncol* 2020;21:73–9.
- Nicholson BD, Hamilton W, Koshiaris C, Oke JL, Hobbs FDR, Aveyard P. The association between unexpected weight loss and cancer diagnosis in primary care: a matched cohort analysis of 65,000 presentations. *Br J Cancer* 2020;122:1848–56.
- Humphrys E, Walter FM, Rubin G, Emery JD, Johnson M, Richards A, et al. Patient symptom experience prior to a diagnosis of oesophageal or gastric cancer: a multi-methods study. *BJGP Open* 2020;4:bjgpopen20X101001.
- Mills K, Birt L, Emery JD, Hall N, Banks J, Johnson M, et al. Understanding symptom appraisal and help-seeking in people with symptoms suggestive of pancreatic cancer: a qualitative study. *BMJ Open* 2017;7:e015682.
- Hamilton W, Walter FM, Rubin G, Neal RD. Improving early diagnosis of symptomatic cancer. *Nat Rev Clin Oncol* 2016;13:740–9.
- Hippisley-Cox J, Coupland C. Symptoms and risk factors to identify women with suspected cancer in primary care: derivation and validation of an algorithm. *Br J Gen Pract* 2013;63:e11–21.
- Hippisley-Cox J, Coupland C. Symptoms and risk factors to identify men with suspected cancer in primary care: derivation and validation of an algorithm. *Br J Gen Pract* 2013;63:e1–10.
- Sharma A, Kandlakunta H, Nagpal SJS, Feng Z, Hoos W, Petersen GM, et al. Model to determine risk of pancreatic cancer in patients with new-onset diabetes. *Gastroenterology* 2018;155:730–9.e3.
- Chari ST, Maitra A, Matrisian LM, Shrader EE, Wu BU, Kambadakone A, et al. Early Detection Initiative: a randomized controlled trial of algorithm-based screening in patients with new onset hyperglycemia and diabetes for early detection of pancreatic ductal adenocarcinoma. *Contemp Clin Trials* 2022;113:106659.
- Li H, Sun D, Cao M, He S, Zheng Y, Yu X, et al. Risk prediction models for esophageal cancer: a systematic review and critical appraisal. *Cancer Med* 2021;10:7265–76.
- Santos R, Coleman HG, Cairnduff V, Kunzmann AT. Clinical prediction models for pancreatic cancer in general and at-risk populations: a systematic review. *Am J Gastroenterol* 2023;118:26–40.
- Gu J, Chen R, Wang SM, Li M, Fan Z, Li X, et al. Prediction models for gastric cancer risk in the general population: a systematic review. *Cancer Prev Res (Phila)* 2022;15:309–18.
- Chen R, Zheng R, Zhou J, Li M, Shao D, Li X, et al. Risk prediction model for esophageal cancer among general population: a systematic review. *Front Public Health* 2021;9:680967.
- Funston G, Hamilton W, Abel G, Crosbie EJ, Rous B, Walter FM. The diagnostic performance of CA125 for the detection of ovarian and non-ovarian cancer in primary care: a population-based cohort study. *PLoS Med* 2020;17:e1003295.
- Mounce LTA, Calitri R, Hamilton W, Rafiq M, Emery J, Giannakeas V, et al. Improving the clinical utility of platelet count for cancer detection in primary care: a cohort study in England, Canada, and Australia. *Cancers (Basel)* 2024; 16:3074.
- Ankus E, Price SJ, Ukoumunne OC, Hamilton W, Bailey SER. Cancer incidence in patients with a high normal platelet count: a cohort study using primary care data. *Fam Pract* 2018;35:671–5.
- Herbert A, Rafiq M, Pham TM, Renzi C, Abel GA, Price S, et al. Predictive values for different cancers and inflammatory bowel disease of 6 common abdominal symptoms among more than 1.9 million primary care patients in the UK: a cohort study. *PLoS Med* 2021;18:e1003708.

31. Rubinstein WS, Patriotis C, Dickherber A, Han PKJ, Katki HA, LeeVan E, et al. Cancer screening with multicancer detection tests: a translational science review. *CA Cancer J Clin* 2024;74:368–82.
32. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71.
33. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015;350:g7594.
34. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–36.
35. Luan Y, Zhong G, Li S, Wu W, Liu X, Zhu D, et al. A panel of seven protein tumour markers for effective and affordable multi-cancer early detection by artificial intelligence: a large-scale and multicentre case-control study. *EClinicalMedicine* 2023;61:102041.
36. Ahmadi N, Mbuagbaw L, Hanna WC, Finley C, Agzarian J, Wen CK, et al. Development of a clinical score to distinguish malignant from benign esophageal disease in an undiagnosed patient population referred to an esophageal diagnostic assessment program. *J Thorac Dis* 2020;12:191–8.
37. Gao Y, Xin L, Lin H, Yao B, Zhang T, Zhou A-J, et al. Machine learning-based automated sponge cytology for screening of oesophageal squamous cell carcinoma and adenocarcinoma of the oesophagogastric junction: a nationwide, multicohort, prospective study. *Lancet Gastroenterol Hepatol* 2023;8:432–45.
38. Holmberg D, Ness-Jensen E, Mattsson F, Lagergren J. Clinical prediction model for tumor progression in Barrett's esophagus. *Surg Endosc* 2019;33:2901–8.
39. In H, Solsky I, Castle PE, Schechter CB, Parides M, Friedmann P, et al. Utilizing cultural and ethnic variables in screening models to identify individuals at high risk for gastric cancer: a pilot study. *Cancer Prev Res (Phila)* 2020;13:687–98.
40. Khademi H, Radmard A-R, Malekzadeh F, Kamangar F, Nasseri-Moghaddam S, Johansson M, et al. Diagnostic accuracy of age and alarm symptoms for upper GI malignancy in patients with dyspepsia in a GI clinic: a 7-year cross-sectional study. *PLoS One* 2012;7:39173.
41. Kurita Y, Kuwahara T, Hara K, Mizuno N, Okuno N, Matsumoto S, et al. Diagnostic ability of artificial intelligence using deep learning analysis of cyst fluid in differentiating malignant from benign pancreatic cystic lesions. *Sci Rep* 2019;9:6893.
42. Manuel-Vázquez A, Balakrishnan A, Agami P, Andersson B, Berrevoet F, Besselink MG, et al. A scoring system for predicting malignancy in intra-ductal papillary mucinous neoplasms of the pancreas: a multicenter EURO-PEAN validation. *Langenbecks Arch Surg* 2022;407:3447–55.
43. Muhammad W, Hart GR, Nartowt B, Farrell JJ, Johung K, Liang Y, et al. Pancreatic cancer prediction through an artificial neural network. *Front Artif Intell* 2019;2:2.
44. Nakatochi M, Lin Y, Ito H, Hara K, Kinoshita F, Kobayashi Y, et al. Prediction model for pancreatic cancer risk in the general Japanese population. *PLoS One* 2018;13:e0203386.
45. Qiu L, Qu X, He J, Cheng L, Zhang R, Sun M, et al. Predictive model for risk of gastric cancer using genetic variants from genome-wide association studies and high-evidence meta-analysis. *Cancer Med* 2020;9:7310–6.
46. Qu W, Zhou Z, Yuan G, Li S, Li J, Chu Q, et al. Is the radiomics-clinical combined model helpful in distinguishing between pancreatic cancer and mass-forming pancreatitis? *Eur J Radiol* 2023;164:110857.
47. Rubenstein JH, Fontaine S, MacDonald PW, Burns JA, Evans RR, Arasim ME. Predicting incident adenocarcinoma of the esophagus or gastric cardia using machine learning of electronic health records. *Gastroenterology* 2023;165:1420–29.e10.
48. Yang M, Zhang XL, Qian W, Yang F, Cai Y, Zhu W, et al. A scoring system for differential diagnosis of benign and malignant pancreatic lesions: establishment and validation. *Chin Acad J Second Mil Med Univ* 2014;35:637–43.
49. Zardab M, Balarajah V, Banerjee A, Stasinou K, Saad A, Imrali A, et al. Differentiating ductal adenocarcinoma of the pancreas from benign conditions using routine health records: a prospective case-control study. *Cancers (Basel)* 2023;15:280.
50. Zhang P, Xu T, Wang J, Li Y, Cai Y, Feng H, et al. Comparison of different risk stratifications for gastric cancer and establishing a simplified risk-scoring model based on the Kyoto classification. *J Gastroenterol Hepatol* 2023;38:1988–97.
51. Zhu S-L, Dong J, Zhang C, Huang Y-B, Pan W. Application of machine learning in the diagnosis of gastric cancer based on noninvasive characteristics. *PLoS One* 2020;15:e0244869.
52. Henriksen SD, Stubbe BE, Madsen PH, Johansen JS, Jensen BV, Hansen CP, et al. Cell-free DNA promoter hypermethylation as a diagnostic marker for pancreatic ductal adenocarcinoma - an external validation study. *Pancreatology* 2021;21:1081–91.
53. Kamran U, King D, Banks M, Nylander D, Shetty S, Hebbar S, et al. Assessment of the role of the Edinburgh dysphagia score in referral triage in a national service evaluation of the urgent suspected upper gastrointestinal cancer pathway. *Aliment Pharmacol Ther* 2022;55:1160–8.
54. Li H, Ding C, Zeng H, Zheng R, Cao M, Ren J, et al. Improved esophageal squamous cell carcinoma screening effectiveness by risk-stratified endoscopic screening: evidence from high-risk areas in China. *Cancer Commun (Lond)* 2021;41:715–25.
55. Park J, Artin MG, Lee KE, Pumpalova YS, Ingram MA, May BL, et al. Deep learning on time series laboratory test results from electronic health records for early detection of pancreatic cancer. *J Biomed Inform* 2022;131:104095.
56. Wang F, Yang Z, Chen XB, Peng Y, Jiang HX, Qin SY. A novel diagnostic model for insulinoma. *Discov Oncol* 2022;13:68.
57. Xiang F, Meng Q-T, Deng J-J, Wang J, Liang X-Y, Liu X-Y, et al. A deep learning model based on contrast-enhanced computed tomography for differential diagnosis of gallbladder carcinoma. *Hepatobiliary Pancreat Dis Int* 2024;23:376–84.
58. Zhu Z, Luo K, Zhang B, Wang G, Guo K, Huang P, et al. Risk factor analysis and construction of prediction models of gallbladder carcinoma in patients with gallstones. *Front Oncol* 2023;13:1037194.
59. Hippisley-Cox J, Coupland C. Identifying patients with suspected gastro-oesophageal cancer in primary care: derivation and validation of an algorithm. *Br J Gen Pract* 2011;61:e707–14.
60. Hippisley-Cox J, Coupland C. Identifying patients with suspected pancreatic cancer in primary care: derivation and validation of an algorithm. *Br J Gen Pract* 2012;62:38–45.
61. Ho KMA, Rosenfeld A, Hogan Á, McBain H, Duku M, Wolfson PB, et al. Development and validation of a multivariable risk factor questionnaire to detect oesophageal cancer in 2-week wait patients. *Clin Res Hepatol Gastroenterol* 2023;47:102087.
62. Khan S, Bhushan B. Machine learning predicts patients with new-onset diabetes at risk of pancreatic cancer. *J Clin Gastroenterol* 2023;58:681–91.
63. Blyuss O, Zaikin A, Cherepanova V, Munblit D, Kiseleva EM, Prytomanova OM, et al. Development of PancRISK, a urine biomarker-based risk score for stratified screening of pancreatic cancer patients. *Br J Cancer* 2020;122:692–6.
64. Cai Q, Zhu C, Yuan Y, Feng Q, Feng Y, Hao Y, et al. Development and validation of a prediction rule for estimating gastric cancer risk in the Chinese high-risk population: a nationwide multicentre study. *Gut* 2019;68:1576–87.
65. Chen Q, Cherry DR, Nalawade V, Qiao EM, Kumar A, Lowy AM, et al. Clinical data prediction model to identify patients with early-stage pancreatic cancer. *JCO Clin Cancer Inform* 2021;5:279–87.
66. Lee H-A, Chen K-W, Hsu C-Y. Prediction model for pancreatic cancer-A population-based study from NHIRD. *Cancers (Basel)* 2022;14:882.
67. Lee D-H, Yoon W, Lee A, Han Y, Byun Y, Kang JS, et al. Multi-biomarker panel prediction model for diagnosis of pancreatic cancer. *J Hepatobiliary Pancreat Sci* 2023;30:122–32.
68. Liu J, Lin S-R, Li ZP, Zhou LY, Xue Y, Yan XE. A novel gastric juice index model for detecting early gastric cancer. *Int J Clin Exp Med* 2017;10:14425–35.
69. Mohammadnezhad K, Sahebi MR, Alatab S, Sadjadi A. Modeling epidemiology data with machine learning technique to detect risk factors for gastric cancer. *J Gastrointest Cancer* 2024;55:287–96.
70. Zhang P-P, Wang T, Li S-Y, Li L, Su X-J, Gu P-Y, et al. Independent risk factors for true malignancy in atypical cytologic diagnostic category in EUS-FNA/FNB of the pancreas: a novel prediction model. *Endosc Ultrasound* 2022;11:208–15.
71. Henriksen SD, Madsen PH, Larsen AC, Johansen MB, Drewes AM, Pedersen IS, et al. Cell-free DNA promoter hypermethylation in plasma as a diagnostic marker for pancreatic adenocarcinoma. *Clin Epigenetics* 2016;8:117.
72. Debernardi S, Blyuss O, Rycyk D, Srivastava K, Jeon CY, Cai H, et al. Urine biomarkers enable pancreatic cancer detection up to 2 years before diagnosis. *Int J Cancer* 2023;152:769–80.

73. Liu M, Zhou R, Liu Z, Guo C, Xu R, Zhou F, et al. Update and validation of a diagnostic model to identify prevalent malignant lesions in esophagus in general population. *EclinicalMedicine* 2022;47:101394.
74. Zhang H, Han W, Jin M, Lai Y, Wang X, Wang J, et al. Establishment and verification of a scoring model for the differential diagnosis of pancreatic cancer and chronic pancreatitis. *Pancreas* 2018;47:459–65.
75. Briggs E, de Kamps M, Hamilton W, Johnson O, McInerney CD, Neal RD. Machine learning for risk prediction of oesophago-gastric cancer in primary care: comparison with existing risk-assessment tools. *Cancers (Basel)* 2022;14:5023.
76. Corona G, Cannizzaro R, Miolo G, Caggiari L, De Zorzi M, Repetto O, et al. Use of metabolomics as a complementary omic approach to implement risk criteria for first-degree relatives of gastric cancer patients. *Int J Mol Sci* 2018;19:750.
77. Etemadi A, Abnet CC, Golozar A, Malekzadeh R, Dawsey SM. Modeling the risk of esophageal squamous cell carcinoma and squamous dysplasia in a high risk area in Iran. *Arch Iran Med* 2012;15:18–21.
78. Liu J, Hu L, Zhou B, Wu C, Cheng Y. Development and validation of a novel model incorporating MRI-based radiomics signature with clinical biomarkers for distinguishing pancreatic carcinoma from mass-forming chronic pancreatitis. *Transl Oncol* 2022;18:101357.
79. Mather MW, Wilson JA, Doona M, Talks BJ, Fullard M, Griffin M, et al. The utility of a symptom model to predict the risk of oesophageal cancer. *Surgeon* 2023;21:119–27.
80. Shen Y, Xie S, Zhao L, Song G, Shao Y, Hao C, et al. Estimating individualized absolute risk for esophageal squamous cell carcinoma: a population-based study in high-risk areas of China. *Front Oncol* 2020;10:598603.
81. Yokoyama T, Yokoyama A, Kumagai Y, Omori T, Kato H, Igaki H, et al. Health risk appraisal models for mass screening of esophageal cancer in Japanese men. *Cancer Epidemiol Biomarkers Prev* 2008;17:2846–54.
82. Ahmed AE, Alzahran FS, Gharawi AM, Alammary SA, Almjimaj FH, Alhusayni FM, et al. Improving risk prediction for pancreatic cancer in symptomatic patients: a Saudi Arabian study. *Cancer Manag Res* 2018;10:4981–6.
83. Cai Q-C, Chen Y, Xiao Y, Zhu W, Xu Q-F, Zhong L, et al. A prediction rule for estimating pancreatic cancer risk in chronic pancreatitis patients with focal pancreatic mass lesions with prior negative EUS-FNA cytology. *Scand J Gastroenterol* 2011;46:464–70.
84. Chang J, Huang Y, Wei L, Ma B, Miao X, Li Y, et al. Risk prediction of esophageal squamous-cell carcinoma with common genetic variants and lifestyle factors in Chinese population. *Carcinogenesis* 2013;34:1782–6.
85. Crouwel F, Meurs-Szojda MM, Klemm-Kropp M, Fockens P, Grasmann ME. The diagnostic yield of open-access endoscopy of the upper gastrointestinal tract in the Netherlands. *Endosc Int Open* 2018;6:E383–94.
86. Dong J, Buas MF, Gharahkhani P, Kendall BJ, Onstad L, Zhao S, et al. Determining risk of Barrett's esophagus and esophageal adenocarcinoma based on epidemiologic factors and genetic variants. *Gastroenterology* 2018;154:1273–81.e3.
87. Lee D-S, Yang H-K, Kim J-W, Yook J-W, Jeon S-H, Kang S-H, et al. Identifying the risk factors through the development of a predictive model for gastric cancer in South Korea. *Cancer Nurs* 2009;32:135–42.
88. Zhou D, Wang J-D, Yang Y, Yu W-L, Zhang Y-J, Quan Z-W. Individualized nomogram improves diagnostic accuracy of stage I-II gallbladder cancer in chronic cholecystitis patients with gallbladder wall thickening. *Hepatobiliary Pancreat Dis Int* 2016;15:180–8.
89. Placido D, Yuan B, Hjaltekin JX, Zheng C, Haue AD, Chmura PJ, et al. A deep learning algorithm to predict risk of pancreatic cancer from disease trajectories. *Nat Med* 2023;29:1113–22.
90. Salvatore M, Beesley LJ, Fritsche LG, Hanauer D, Shi X, Mondul AM, et al. Phenotype risk scores (PheRS) for pancreatic cancer using time-stamped electronic health record data: discovery and validation in two large biobanks. *J Biomed Inform* 2021;113:103652.
91. Huang RJ, Kwon NS-E, Tomizawa Y, Choi AY, Hernandez-Boussard T, Hwang JH. A comparison of logistic regression against machine learning algorithms for gastric cancer risk prediction within real-world clinical data streams. *JCO Clin Cancer Inform* 2022;6:2200039.
92. Kapoor N, Bassi A, Sturgess R, Bodger K. Predictive value of alarm features in a rapid access upper gastrointestinal cancer service. *Gut* 2005;54:40–5.
93. Shen X, Lu D, Xu X, Wang J, Wu J, Yan S, et al. A novel distinguishing system for the diagnosis of malignant pancreatic cystic neoplasm. *Eur J Radiol* 2013;82:e648–54.
94. Tao W, Wang H-X, Guo Y-F, Yang L, Li P. Establish a scoring model for high-risk population of gastric cancer and study on the pattern of opportunistic screening. *Gastroenterol Res Pract* 2020;2020:5609623.
95. Zhou R, Zheng H, Liu M, Liu Z, Guo C, Tian H, et al. Development and validation of a questionnaire-based risk scoring system to identify individuals at high risk for gastric cancer in Chinese populations. *Chin J Cancer Res* 2021;33:649–58.
96. Yokoyama A, Oda J, Iriguchi Y, Kumagai Y, Okamura Y, Matsuoka M, et al. A health-risk appraisal model and endoscopic mass screening for esophageal cancer in Japanese men. *Dis Esophagus* 2013;26:148–53.
97. Zhou Q, Chen Y, Pan J, Zhou L, Lin J. Application of a novel scoring system for gastric cancer opportunistic screening in hospital visits. *BMC Gastroenterol* 2022;22:223.
98. Hu Y, Bao H, Jin H, Zhao J, Xu Y, Huang X, et al. Performance evaluation of four prediction models for risk stratification in gastric cancer screening among a high-risk population in China. *Gastric Cancer* 2021;24:1194–202.
99. NG12 NICE. Suspected cancer: recognition and referral 2023. [cited 2025 Mar 4]. Available from: <https://www.nice.org.uk/guidance/ng12>.
100. Archer S, Donoso FS, Carver T, Lin J, Cunningham AP, Ficorella L, et al. Exploring the barriers to and facilitators of implementing CanRisk in primary care: a qualitative thematic framework analysis. *Br J Gen Pract* 2023;73:586–96.
101. Blackford AL, Canto MI, Dbouk M, Hruban RH, Katona BW, Chak A, et al. Pancreatic cancer surveillance and survival of high-risk individuals. *JAMA Oncol* 2024;10:1087–96.
102. Collins GS, Dhiman P, Ma J, Schlüssel MM, Archer L, Van Calster B, et al. Evaluation of clinical prediction models (part 1): from development to external validation. *BMJ* 2024;384:e074819.
103. Rubin G, Walter FM, Emery J, Hamilton W, Hoare Z, Howse J, et al. Electronic clinical decision support tool for assessing stomach symptoms in primary care (ECASS): a feasibility study. *BMJ Open* 2021;11:e041795.
104. Akanuwe JNA, Black S, Owen S, Siriwardena AN. Barriers and facilitators to implementing a cancer risk assessment tool (QCancer) in primary care: a qualitative study. *Prim Health Care Res Dev* 2021;22:e51.
105. Chiang PP-C, Glance D, Walker J, Walter FM, Emery JD. Implementing a QCancer risk tool into general practice consultations: an exploratory study using simulated consultations with Australian general practitioners. *Br J Cancer* 2015;112(Suppl 1):S77–83.
106. Black GB, Machen S, Parker-Deeks S, Cronin A, Chung D. Using an electronic safety netting tool designed to improve safety with respect to cancer referral in primary care: a qualitative service evaluation using rapid appraisal methods. *BMJ Open Qual* 2023;12:e002354.
107. Dikomitis L, Green T, Macleod U. Embedding electronic decision-support tools for suspected cancer in primary care: a qualitative study of GPs' experiences. *Prim Health Care Res Dev* 2015;16:548–55.
108. Frisinger A, Papachristou P. The voice of healthcare: introducing digital decision support systems into clinical practice – a qualitative study. *BMC Prim Care* 2023;24:67.
109. Helenason J, Ekström C, Falk M, Papachristou P. Exploring the feasibility of an artificial intelligence based clinical decision support system for cutaneous melanoma detection in primary care – a mixed method study. *Scand J Prim Health Care* 2024;42:51–60.
110. Chima S, Martinez-Gutierrez J, Hunter B, Manski-Nankervis J-A, Emery J. Optimization of a quality improvement tool for cancer diagnosis in primary care: qualitative study. *JMIR Form Res* 2022;6:e39277.
111. Buck C, Doctor E, Hennrich J, Jöhnk J, Eymann T. General practitioners' attitudes toward artificial intelligence-enabled systems: interview study. *J Med Internet Res* 2022;24:e28916.
112. Ho S, Doig GS, Ly A. Attitudes of optometrists towards artificial intelligence for the diagnosis of retinal disease: a cross-sectional mail-out survey. *Ophthalmic Physiol Opt* 2022;42:1170–9.
113. Derksen C, Scott S, Akbar AB, Parma AVE, Saunders T, Round T, et al. Computerised clinical decision support systems (CDSS) for the detection of disease in primary care: systematic review and recommendations for disease detection tools. PROSPERO. Southampton (UK): National Institute for Health and Care Research; 2024. [cited 2025 Mar 4]. Available from: <https://www.crd.york.ac.uk/PROSPERO/view/CRD42024517054>.
114. Guerra CE, Sharma PV, Castillo BS. Multi-cancer early detection: the new frontier in cancer early detection. *Annu Rev Med* 2024;75:67–81.