

RESEARCH ARTICLE

Open Access



# Development, validation and effectiveness of diagnostic prediction tools for colorectal cancer in primary care: a systematic review

Bogdan Grigore<sup>1\*</sup> , Ruth Lewis<sup>2</sup>, Jaime Peters<sup>1</sup>, Sophie Robinson<sup>3</sup> and Christopher J. Hyde<sup>1</sup>

## Abstract

**Background:** Tools based on diagnostic prediction models are available to help general practitioners (GP) diagnose colorectal cancer. It is unclear how well they perform and whether they lead to increased or quicker diagnoses and ultimately impact on patient quality of life and/or survival. The aim of this systematic review is to evaluate the development, validation, effectiveness, and cost-effectiveness, of cancer diagnostic tools for colorectal cancer in primary care.

**Methods:** Electronic databases including Medline and Web of Science were searched in May 2017 (updated October 2019). Two reviewers independently screened titles, abstracts and full-texts. Studies were included if they reported the development, validation or accuracy of a prediction model, or assessed the effectiveness or cost-effectiveness of diagnostic tools based on prediction models to aid GP decision-making for symptomatic patients presenting with features potentially indicative of colorectal cancer. Data extraction and risk of bias were completed by one reviewer and checked by a second. A narrative synthesis was conducted.

**Results:** Eleven thousand one hundred thirteen records were screened and 23 studies met the inclusion criteria. Twenty-studies reported on the development, validation and/or accuracy of 13 prediction models: eight for colorectal cancer, five for cancer areas/types that include colorectal cancer. The Qcancer models were generally the best performing.

Three impact studies met the inclusion criteria. Two (an RCT and a pre-post study) assessed tools based on the RAT prediction model. The third study looked at the impact of GP practices having access to RAT or Qcancer. Although the pre-post study reported a positive impact of the tools on outcomes, the results of the RCT and cross-sectional survey found no evidence that use of, or access to, the tools was associated with better outcomes. No study evaluated cost effectiveness.

**Conclusions:** Many prediction models have been developed but none have been fully validated. Evidence demonstrating improved patient outcome of introducing the tools is the main deficiency and is essential given the imperfect classification achieved by all tools. This need is emphasised by the equivocal results of the small number of impact studies done so far.

**Keywords:** Cancer, Primary care, Diagnostic prediction models

\* Correspondence: [b.grigore@exeter.ac.uk](mailto:b.grigore@exeter.ac.uk)

<sup>1</sup>Exeter Test Group, College of Medicine and Health, University of Exeter Medical School, Exeter, UK

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Colorectal cancer is the third most frequent cancer and the second leading cause of cancer-related death in the world [1]. In 2014–2016 there were 42,042 new cases of colorectal cancer in the UK, with 57% of people with colorectal cancer surviving for 10 years or more [2].

Research suggests that cancer prognosis can be improved by reducing the time to diagnosis [3], as earlier diagnosis is associated with earlier stage at diagnosis [4], and earlier treatment is associated with improved survival [5]. Reducing time to diagnosis also has the potential to reduce presentation via emergency admissions, and prevent the poorer survival associated with that route of diagnosis [6]. A national cancer screening programme exists in the National Health Service (NHS) for colorectal cancer, and the National Awareness and Early Diagnosis Initiative (NAEDI) (to increase public awareness on the signs and symptoms of cancer [7]) is intended to improve early diagnosis. However, as many individuals go through primary care as a route for diagnosis [6], so efforts here could improve cancer survival.

Cancer diagnosis in primary care is not straightforward. Symptoms of cancer are commonly seen but mostly have non-cancer origins [8]. Of those individuals referred from primary care via the two-week wait (2WW) referrals for suspected colorectal cancer in areas of England, approximately 5–8% were ultimately diagnosed with cancer [9, 10]. The type and presence of symptoms can vary greatly [11] and it is not surprising that patients can have multiple general practitioner (GP) consultations before being referred, especially for those cancers that have less well-known signs and symptoms [12]. Thus, tools to help improve cancer diagnosis in primary care have great potential to impact on diagnoses and subsequent treatment options, leading to better outcomes for patients.

Diagnostic prediction models combine multiple predictors, such as symptoms and patient characteristics, to obtain the risk of the presence or absence of a disease within an individual patient [13, 14]. These prediction models can then be used to develop diagnostic tools (such as a website risk calculator, or mouse mat containing estimates of risk depending on features) to assist doctors in estimating probabilities and potentially influence their decision making [14]. To evaluate diagnostic prediction models, there are three important stages, or types of studies: prediction model development, prediction model validation, and assessment of the impact of prediction models in practice (generally implemented as diagnostic tools). The first two are often conducted as part of the same study, and are generally evaluated using a single cohort design. These types of studies are commonly found in the diagnostic prediction literature, with some studies also reporting results of an external

validation [15]. To assess the impact of the prediction model (the third stage), comparative studies are required to evaluate the ability of the tool to guide patient management. However, very few diagnostic prediction models that are developed go on to be evaluated for their clinical impact [15] or cost-effectiveness.

Tools currently available to GPs in the UK to help cancer diagnosis, beyond the National Institute for Health and Care Excellence (NICE) guidelines for suspected cancer referral [8], are based on diagnostic prediction models, and are integrated into GP software systems.

- 1 The Risk Assessment Tool (RAT) developed by Hamilton and colleagues which provides estimates of cancer risk for 17 cancers based on symptoms alone is integrated into Vision (INPS), and
- 2 The Qcancer tool, which estimates the risk of 11 cancers based on symptoms and patient characteristics, and overall cancer risk in males and females, is integrated into EMIS Web.

There is recent evidence that these tools are being used in primary care [16], however it is unclear whether these tools impact on GP decision-making, and ultimately on patient outcomes.

Systematic reviews have looked at the use of prediction models for colorectal cancer in primary and secondary care [17]. However, more research in the primary care setting had been published for colorectal cancer since, so we sought to systematically review this evidence. The aim of our review was to identify reports on the development, validation or accuracy of prediction models, as well as evidence evaluating the impact (i.e. effectiveness or cost-effectiveness) of symptom-based diagnostic tools that could be used to inform colorectal cancer diagnosis decision-making in primary care.

## Methods

This systematic review was conducted as part of a wider programme considering risk assessment tools for any cancer site [18]. Protocols relevant to the systematic review described here were registered on PROSPERO (CRD42017068373, CRD42017068375).

The systematic review was conducted in accordance with good practice guidelines [19] and is reported here in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [20].

## Search strategy

Bibliographic searches of relevant databases (Medline, Medline in Process, Embase, Cochrane, Web of Science), were conducted in May 2017 and updated in October 2019.

The search strategies were developed by an information specialist (SR) and comprised terms for cancer, terms for

primary care, terms for decision support tools and terms for diagnosis (see supplementary Table S1). No date, language or other limits were used. Search filters for clinical prediction models were investigated but none were thought to be fully tested or reliable. A balance was sought between sensitivity of search results and volume of papers to screen. As the search strategies were originally developed to identify reports related to prediction models for any cancer site [18], no cancer site specific terms were used. Instead, we retrospectively excluded non-colorectal cancer studies from the current systematic review.

The search results were exported to Endnote X7 (Thomson Reuters, NY, USA) and de-duplicated using automatic and manual checking.

Additional searches were conducted using Scopus (Elsevier) on the references, as well as any citations of the items included after full-text screening, in order to identify additional relevant studies. Searches were also conducted for identified named tools (QCancer, RAT, CAPER, Bristol-Birmingham equation) in order to ensure search results were sufficiently comprehensive.

#### **Inclusion and exclusion criteria**

Diagnostic prediction models are defined as multivariate statistical models that predict the probability or risk that a patient currently has cancer based on a combination of known features of that patient, such as symptoms, signs, test results and patient characteristics [21]. Symptoms could be self-reported by the patient, or prompted by physician's questioning. Signs and test results are identified within primary care via routine testing (such as full blood count, urine dipstick testing, clinical signs), as are patient characteristics (socio-demographic variables, personal and family history). Studies that simply looked at 'red-flag symptoms' or symptom lists and (weighted) scores that did not provide a numerical risk of current cancer were excluded. Models developed with secondary care data (i.e. referred patients) were only included if an attempt was made to validate the models with primary care data.

Inclusion and exclusion criteria are presented in Table 1.

#### **Selection of studies**

Titles and abstracts were screened for relevance independently (by BG and RL), and any disagreements were resolved by consensus. Pilot screening was undertaken for the first 100 hits to ensure both reviewers were interpreting the inclusion and exclusion criteria in the same way. Articles retained were obtained in full and further screened independently by the two reviewers. For any disagreements that were not resolved, a third reviewer (CH) made the final decision.

The development and validation aspects of particular prediction models were often reported in multiple

studies (e.g. the development and internal validation of the Q cancer prediction model was presented in one paper by Hippisley-Cox and colleagues, 2012 [22] and the external validation in a separate paper (Collins and colleagues, 2012 [23]) All studies related to the same specific prediction model were collated regardless of whether they refer to the development, validation and/or impact of that tool.

#### **Data extraction**

To extract relevant data from each included study, standardised data extraction forms were used that evolved following piloting and discussion among reviewers. One reviewer (BG) extracted the data, which was checked by a second reviewer (RL). The following data were extracted from all study types: included cancer type(s), study design, country, sample size, participant recruitment (with inclusion and exclusion criteria) and participant characteristics. For studies reporting on the development and/or validation of prediction models an adaptation of the CHARMS checklist (CHECKlist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies) [24] was used to extract additional relevant data, including data source, number of participants with specific cancer, features of the model (what symptoms, test results, patient demographics etc. are included), how features are defined and measured, definition of primary and secondary outcomes, how and when outcomes are assessed, main results (including model performance, validation and estimates of risk), features included in final model. For studies reporting the impact of tools based on prediction models additional items extracted included characteristics of the tool (including whether based on symptoms alone or other features in addition to symptoms), definition of outcomes, main results including confidence intervals, and subgroup analyses, where available.

#### **Risk of bias assessment**

Risk of bias of studies reporting the development and/or validation of prediction models was assessed with the PROBAST [25] (Prediction model Risk of Bias ASsessment Tool) checklist. The derived checklist assesses the risk of bias and applicability of prediction-modelling studies on 5 domains: participant selection, predictors, outcome, sample size and missing data, analysis.

For studies reporting on the impact of tools based on decision models a risk-of-bias form based on the Cochrane EPOC (Effective Practice and Organisation of Care) group recommendations [26] was used. All risk of bias assessments were conducted by one reviewer (BG) and checked by a second reviewer (RL).

**Table 1** Inclusion and exclusion criteria

|                         |  |
|-------------------------|--|
| <b>Population</b>       | Included: adult symptomatic patients (with symptoms being indicative of cancer) presenting at primary care or patients referred with symptoms indicative of cancer<br><br>Excluded: asymptomatic patients (screening population).  |
| <b>Technology</b>       | Included: Diagnostic prediction models, based on 2 or more features <sup>a</sup> , that estimate the risk of prevalent but undiagnosed colorectal cancer.<br><br>Excluded: prognostic or screening prediction models<br>Statistical tools that estimate the probability of developing cancer over a defined period of time.<br>Prediction models that did not include colorectal cancer.   |
| <b>Setting</b>          | Included: primary care<br><br>Excluded: secondary care; on-line tools developed for use by the general population  |
| <b>Study design</b>     | Included:<br>- any design for the <b>development, validation or accuracy</b> of diagnostic prediction models (as defined under 'Technology');<br>- comparative studies of diagnostic tools that assessed <b>impact</b> in clinical practice (Randomised controlled trials, controlled before-after, and interrupted time-series;<br>studies analysing national trends in cancer diagnosis before and after diagnostic tools became available)<br><br>Excluded: uncontrolled studies reporting qualitative data   |
| <b>Comparison</b>       | Usual care or the use of another diagnostic tool   |
| <b>Outcomes</b>         | <b>For studies reporting development, validation and/or accuracy of prediction models:</b><br>Estimates of the risk of being diagnosed with cancer (e.g. ORs, HRs)<br>AND/OR<br>Any details on the development, validation or accuracy of the tool:<br>• <i>Model development</i> : method; assumptions; predictors; shrinkage; coefficient weighting<br>• <i>Model evaluation (validation)</i><br>• <i>Assessing (quantifying) model performance</i> : discrimination (ability to discriminate participants with or without the outcome, e.g. area under the ROC curve); calibration (agreement between predicted and observed outcome); overall performance (for discrimination and calibration, e.g. R2); classification (e.g. sensitivity, specificity, predictive values)<br><b>For studies reporting evaluations of the impact of tools:</b><br><i>Primary outcomes</i><br>- patient-related outcome measures (including the number of cancer diagnoses, time to cancer diagnosis, stage of cancer at diagnosis, resection rates, patient health-related quality of life, other patient-reported outcome measures);<br>- survival;<br>- economic outcome measures (resource use, cost per diagnosis), cost per QALY;<br><i>Secondary outcomes</i><br>- referral patterns.<br><br>Exclude: models that report the risk of survival (or stage at diagnosis etc.) |
| <b>Publication type</b> | Included: Published in full and in English<br><br>Excluded: commentaries, letters  |

*Abbreviations:* HR Hazard ratio, N/A Not applicable, OR Odds ratio, QALY Quality-adjusted life year, ROC Receiver operating characteristic  
Note: <sup>a</sup> Features include symptoms and other information, such as elicited signs, patient characteristics and test results

## Synthesis

Owing to the heterogeneity between included studies a narrative review of the studies was conducted.

## Results

### Studies identified

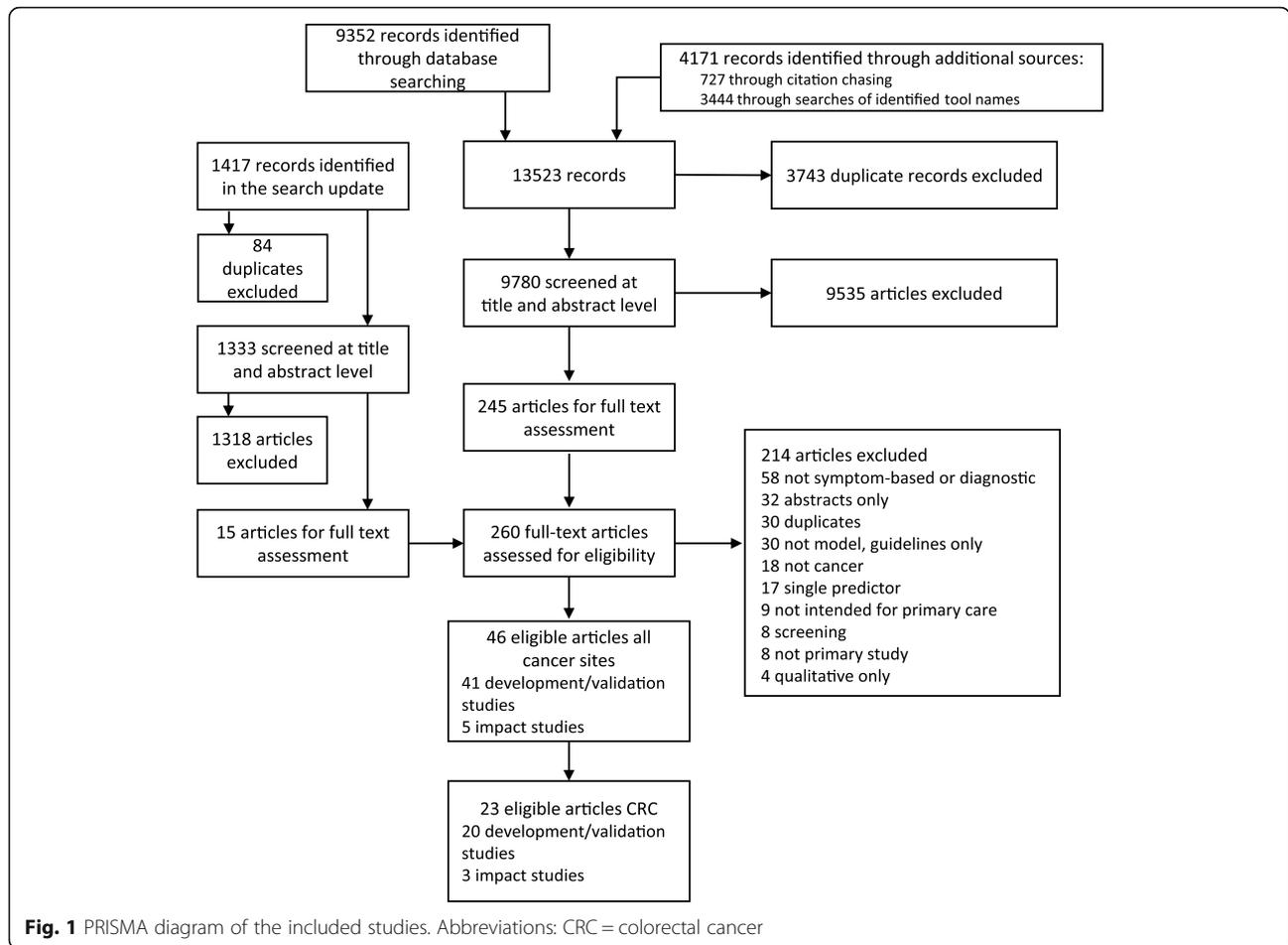
Search phrases were finalised and searches were run in May 2017. A total of 9352 records were obtained through database searching. Additional reference and citation searches on tool names resulted in another 4171 records. After de-duplication, 9780 records were obtained. The database searches were updated in October 2019, and resulted in 2254 additional new records (after de-duplication). After screening the title and abstracts of these records independently by two reviewers, 260 records were retained for full text screening.

We identified two systematic reviews. Scanning their reference list led to the inclusion of two additional studies not found in the database search. One systematic review also included validation of models [27]. In the end, 23 records were identified that were relevant for colorectal cancer (Fig. 1).

Discussions with collaborators led to the identification of relevant grey literature, but no such studies were deemed eligible for inclusion.

### Development/validation studies

Elias and colleagues (2017) [27] aimed to identify and validate published diagnostic models to safely reduce unnecessary endoscopy referrals for colorectal cancer. A systematic review of the literature was undertaken up until 2015 and identified models were validated using a



cross-sectional Dutch dataset referred to as CEDAR ( $n = 810$ ). The definition of model used by Elias and colleagues is very broad and includes guidelines and weighted scores. Therefore, although Elias and colleagues identified 18 models, only four are relevant to our review: Fijten and colleagues (1995) [28] and Marshall and colleagues (2011) [29] which were identified from our searches, while Muris and colleagues (1995) [30] and Nørrelund and colleagues (1996) [31] are new inclusions. Due to the fact that Elias and colleagues attempted to validate the models they found, their validation of these four models is included in the results below.

Of the 20 included model development or validation studies, 17 report on the development (with some also reporting on validation) of models, four only report model validation.

**Prediction models**

The included studies (excluding the validation by Elias and colleagues [27]) reported on 13 different prediction models. Eight models are specifically for colorectal cancer: the Bristol-Birmingham equation (Marshall [29]), a

Dutch model (Fijten [28]), a machine learning algorithm (Kop [32]), a Danish model (Nørrelund [31]), Qcancer (Hippisley-Cox [22]), RAT 2005 (Hamilton [33]), RAT 2009 (Hamilton [34]) and RAT 2017 (Stapley [35]). One model relates to metastatic cancer (RAT, Hamilton [36]), and the remaining four models cover multiple cancer sites which include colorectal cancer: Qcancer for males (Hippisley-Cox [37]), Qcancer for females (Hippisley-Cox [38]), a model for abdominal complaints (Muris [30]), and a model for abdominal cancers (Holtedahll 2018 [39]). Elias [27] and Collins [23] reported on the validation of one or more of the above models.

Table 2 provides a brief description of the models, their stages of development, the cancer sites covered (colorectal cancer-specific or other) and study designs.

The risk prediction models referred to as RATs [33, 35, 36, 43, 44] were designed to be used with patients presenting to primary care with “low-risk-but-not-no-risk symptoms” [45]. Early versions of RATs were developed using case-control data from Devon, UK as part of the CAPER studies [34]. Later models were derived using UK-wide primary care data – the Clinical Practice

**Table 2** Summary of the prediction models, their stages of development, the cancer sites covered and study designs

| Prediction model            | Number and category of descriptors  | Stage of development | Study design               | Country         | Population   | Source  |
|-----------------------------|---|----------------------|----------------------------|-----------------|--|---|
| <b>Colorectal cancer</b>    |   |                      |                            |                 |  |   |
| Bristol-Birmingham equation | 8 Symptoms, Test results  | External validation  | Retrospective Case-control | UK              | Derivation cohort: THIN<br>Validation cohort: CAPER  | Marshall 2011 [29]  |
|                             |   | External validation  | Prospective cohort         | The Netherlands | CEDAR study: Patients referred to endoscopy centres by participating Dutch primary care practices. 2009–2012   | Elias 2017 [27]   |
| Netherlands model           | 3 Symptoms, Patient demographics  | Apparent performance | Prospective cohort         | The Netherlands | 290 consecutive patients with rectal bleeding presenting to 83 GPs in Limburg (Netherlands) September 1988 to April 1990<br>Predictors: Questionnaires completed by GPs and patients, and laboratory test results. | Fijten 1995 [28]  |
|                             |   | External validation  | Prospective cohort         | UK              | patients referred from primary care with colorectal symptoms over a 3-yr period to the Leighton Hospital, Crewe, Cheshire, UK  | Hodder 2005 [40]  |
|                             |   | External validation  | Prospective cohort         | Netherlands     | CEDAR study: Patients referred to endoscopy centres by participating Dutch primary care practices. 2009–2012   | Elias 2017 [27]   |
| Machine learning algorithm  | Numerous models are reported<br>Patient demographics, Symptoms, Medical history, Test results | Apparent performance | Case-control               | The Netherlands | anonymised electronic records from two GP database systems from the Utrecht region, Netherlands, between 01 and 07-2006 and 31-12-2011   | Kop 2015 [41];<br>Kop 2016 [32];<br>Hoogendoorn 2015 [42] |
| Danish model                | 2 Patient demographics<br>Symptoms  | Apparent performance | Prospective cohort         | Denmark         | Patients presenting to GPs with first episode of rectal bleeding.<br>Study 1: 750 GPs 1989–1991<br>Study 2: 450 GPs 1991–1992  | Nørrelund 1996 [31]                                       |
|                             |   | External validation  | Prospective cohort         | The Netherlands | CEDAR study: Patients referred to endoscopy centres by participating Dutch primary care practices. 2009–2012   | Elias 2017 [27]   |
| Qcancer                     | 6 (females)<br>7 (males)<br>Symptoms, Medical history, Test results                           | Internal validation  | open Prospective cohort    | UK              | QResearch database   | Hippisley-Cox 2012c [22]                                  |
|                             |   | External validation  | Prospective cohort         | UK              | THIN database  | Collins 2012 [23]   |
| RAT (2005)                  | 10 Symptoms, Test results   | Apparent performance | Case-control               | UK              | Patients attending all 21 general practices in Exeter, Devon, UK<br>Cases identified from the cancer registry at the Royal Devon and Exeter Hospital   | Hamilton 2005 [33]  |

**Table 2** Summary of the prediction models, their stages of development, the cancer sites covered and study designs (Continued)

| Prediction model                                 | Number and category of descriptors  | Stage of development | Study design            | Country   | Population   | Source                   |
|--|---|----------------------|-------------------------|---|--|--------------------------|
|  |   | External validation  | Prospective cohort      | The Netherlands   | CEDAR study: Patients referred to endoscopy centres by participating Dutch primary care practices. 2009–2012 | Elias 2017 [27]          |
| RAT (2009)                                       | 8 Symptoms, Test results  | Apparent performance | Case-control            | UK  | THIN database  | Hamilton 2009 [43]       |
| RAT (bowel)                                      | 10 Symptoms, Test results   | Apparent performance | Case-control            | UK  | GPRD (currently called the CPRD)   | Stapley 2017 [35]        |
| <b>Metastatic cancer</b>                         |   |                      |                         |   |  |                          |
| RAT  | 7 Symptoms, Test results  | Apparent performance | Case-control            | UK  | Patients attending 11 general practices in Devon, UK   | Hamilton 2015 [36]       |
| <b>Multiple cancer sites</b>                     |   |                      |                         |   |  |                          |
| QCancer (female)                                 | 7 (uterine)<br>10 (breast, blood)<br>11 (ovarian, renal)<br>12 (cervical)<br>13 (colorectal, gastro-oesophageal)<br>14 (pancreatic)<br>15 (lung)<br>22 (other cancers)<br>Medical history, Symptoms, Test results, Patient demographics | Internal validation  | Open prospective cohort | UK  | QResearch database   | Hippisley-Cox 2013 [38]  |
| QCancer (male)                                   | 3 (testicular)<br>8 (renal tract)<br>12 (colorectal)<br>13 (gastro-oesophageal)<br>14 (prostate, blood)<br>15 (pancreatic)<br>17 (lung)<br>20 (other cancers)<br>Medical history, Symptoms, Test results, Patient demographics          | Internal validation  | Open prospective cohort | UK  | QResearch database   | Hippisley-Cox 2013b [37] |
| Muris abdominal complaints model                 | 5 Symptoms<br>Patient demographics<br>Test results  | Apparent performance | Prospective cohort      | The Netherlands   | Patients presenting to GPs for new abdominal complaints. 1989  | Muris 1995 [30]          |
| (Netherlands)                                    |   | External validation  | Prospective cohort      | The Netherlands   | CEDAR study: Patients referred to endoscopy centres by participating Dutch primary care practices. 2009–2012 | Elias 2017 [27]          |
| Abdominal model, Holtedahl and colleagues (2018) | 4 Symptoms, Patient demographics  | Apparent performance | Prospective cohort      | Norway, Denmark, Sweden, Scotland, Belgium, Netherlands | GP records from the participating countries  | Holtedahl, 2018 [39]     |

**Abbreviation:** RAT(s) Risk assessment tool(s)

Research Datalink (formerly General Practice Research Database) [35, 44, 46–51], and The Health Improvement Network (THIN) database [43, 52]. In addition to the

models identified in this systematic review as relevant to colorectal cancer, RATs exist for the following cancer sites: lung, ovarian, kidney, bladder, pancreas, breast,

uterine, brain, prostate, Hodgkin lymphoma, non-Hodgkin lymphoma and multiple myeloma. The RATs are available as prints on common office objects (e.g. mousepads) and are integrated into general practitioner software in the form of the electronic Cancer Decision Support (eCDS). Regardless of the format, they provide risk estimates for patients with single symptoms of possible cancer, pairs of symptoms and repeat attendances with the same symptoms. Elias used a Dutch dataset to externally validate the 2005 colorectal version of RATs [27]. No other RAT was externally validated.

The Qcancer series of models can be used both in symptomatic (diagnostic models) and asymptomatic (prognostic models) patients [53]. Qcancer was developed in the QRESEARCH database, a large database comprising over 12 million anonymised health records from 602 general practices throughout the United Kingdom using the EMIS (Egton Medical Information Systems) computer system. Initially, several models were developed for each cancer type in symptomatic populations, in addition to colorectal: lung, renal, gastro-oesophageal, pancreatic and ovarian cancer. An updated approach incorporates multiple risk factors and symptoms into one model for males and one model for females to predict cancer risk. Most of these models have been externally validated in UK-wide populations (e.g. THIN database [54]). Qcancer is available as an online calculator ([www.qcancer.org](http://www.qcancer.org)), which provides estimates of absolute risk of any cancer with a breakdown of type of cancer based on both risk factors such as age, gender and family history, which increase the likelihood of cancer, and risk markers such as haemoptysis or features, usually symptoms (e.g. weight loss), suggesting that cancer is already present.

Marshall and colleagues (2011) used data from the THIN dataset (>40,000 participants) to construct a model for colorectal cancer, known as the Bristol-Birmingham equation [29]. The model was validated by Marshall et al. using the UK CAPER dataset and was also validated by Elias et al. (26) in a Dutch population. Data from 290 patients presenting to GPs in the Netherlands with rectal bleeding (from 1988 to 1990) were used by Fijten and colleagues (1995) [28] to develop a prediction model for colorectal cancer (Netherlands model). The Netherlands model was validated by Hodder and colleagues (2005) [40] using secondary care data from the UK, and by Elias and colleagues (2017) [27] using a Dutch dataset. Kop and colleagues (2015) [32, 41, 42] used a machine learning algorithm to develop a prediction model for colorectal cancer using electronic records of almost 220,000 patients from two GP practices in the Netherlands. We found no external validation of this model. A Danish colorectal model [31] has also been developed for use in primary care, this was externally validated using a Dutch dataset by Elias and colleagues (2017) [27].

Muris and colleagues (1995) [30] developed a model using data from the Netherlands to predict multiple cancers related to abdominal complaints, which was externally validated by Elias [27].

Holtedahl and colleagues (2018) [39] detail the development of a prediction model for abdominal cancers. These are defined as all cancers of the digestive organs, female genital organs and urinary organs (including testis). Data on 61,802 patients, recorded during GP consultations over a 10 day period from Norway, Denmark, Sweden, Scotland, Belgium, and the Netherlands, were used to develop the model. No validation of the model was identified.

The models are in various stages of development. A total of 5 models (or versions of models) have only assessed apparent performance [35, 36, 39, 41, 43], two models have been internally validated (Qcancer for males and Qcancer for females), one model was updated as a result of using a different data source [43]. One of the four Qcancer versions [22], one RAT version [33] and four of the other prediction models [28–31] have been externally validated, the highest level of evidence identified in this systematic review. Apart from the two Qcancer versions, which were externally validated by Collins and Altman, all other external validations were conducted by Elias et al. [27]. This was a systematic review which used a cross-sectional Dutch dataset referred to as CEDAR ( $n = 810$ ) to validate the models they identified.

All of the models were developed in primary care settings in Europe. Only five models were not derived from UK-only data: Fijten and colleagues (1995) [28], Kop and colleagues (2015) [41], and Muris and colleagues (1995) [30] were developed in the Netherlands, Nørrelund and colleagues (1996) [31] was developed in Denmark, and Holtedahl which used data from Norway, Denmark, Sweden, Scotland, Belgium and the Netherlands. For those models having been externally validated, most were validated in the country in which it was developed except for: the validation [40] of the Netherlands colorectal cancer model [28] in a UK population, the validation of the Danish colorectal cancer [31] in a Dutch population [27] and the validation of the colorectal version of RATs (UK) [33] in a Dutch population [27].

### **Critical appraisal**

The assessment of risk of bias is summarised in Table 3, and given in more detail in supplementary Table S3. Note that for the RATs and Qcancer models, only one entry each is shown as all versions of the RAT or Qcancer model scored the same for each aspect of the risk of bias tool used. Qcancer development and validation studies were judged to be of low risk of bias. For the RAT development studies, there is uncertainty as to the

**Table 3** Risk of bias assessment for the included model development/validation studies

| Model (author of first version)  | Stage of development covered                           | I. Participant selection <sup>a</sup> | II. Predictors <sup>a</sup> | III. Outcome <sup>a</sup> | IV. Sample size and participant flow <sup>a</sup> | V. Analysis <sup>a</sup> |
|--|--|---------------------------------------|-----------------------------|---------------------------|---|--------------------------|
| RAT (Hamilton) series of models for colorectal and meta-static cancer [33, 36, 43]                                       |  |                                       |                             |                           |   |                          |
|  | Apparent performance                                   | ✓                                     | ?                           | ✓                         | ?   | x                        |
|  | External validation (colorectal only) [27]             | ✓                                     | ✓                           | ✓                         | ?   | ?                        |
| Qcancer (Hippisley-Cox) series of models for <b>colorectal and multiple sites for females and males</b> [22, 23, 37, 38] |  |                                       |                             |                           |   |                          |
|  | Internal validation                                    | ✓                                     | ✓                           | ✓                         | ✓   | ✓                        |
|  | External validation (colorectal only) [23]             | ✓                                     | ✓                           | ✓                         | ✓   | ✓                        |
| Bristol-Birmingham (Marshall) [29] model for <b>colorectal cancer</b>  |  |                                       |                             |                           |   |                          |
|  | External validation                                    | ✓                                     | ?                           | ✓                         | ?   | ✓                        |
|  | External validation (Elias and colleagues, 2017) [27]  | ✓                                     | ✓                           | ✓                         | ?   | ?                        |
| Netherlands' (Fitjen 1995 [28]) model for <b>colorectal cancer</b>   |  |                                       |                             |                           |   |                          |
|  | Apparent performance                                   | x                                     | ✓                           | ✓                         | ?   | x                        |
|  | External validation (Hodder and colleagues, 2005) [40] | x                                     | ?                           | x                         | ✓   | ?                        |
|  | External validation (Elias and colleagues, 2017) [27]  | ✓                                     | ✓                           | ✓                         | ?   | ?                        |
| Netherlands' (Kop) [32] 'machine learning' for <b>colorectal cancer</b>  |  |                                       |                             |                           |   |                          |
|  | Apparent performance                                   | ✓                                     | ?                           | ✓                         | ?   | ?                        |
| Danish (Nørrelund 1996 [31]) model for <b>colorectal cancer</b>  |  |                                       |                             |                           |   |                          |
|  | Apparent performance                                   | ✓                                     | ?                           | ✓                         | ?   | x                        |
|  | External validation (Elias and colleagues, 2017) [27]  | ✓                                     | ✓                           | ✓                         | ?   | ?                        |
| Netherlands' (Muris 1995 [30]) model for <b>abdominal complaints</b>   |  |                                       |                             |                           |   |                          |
|  | Apparent performance                                   | ?                                     | ✓                           | ✓                         | ?   | x                        |
|  | External validation (Elias and colleagues, 2017) [27]  | ✓                                     | ✓                           | ✓                         | ?   | ?                        |
| Prediction model for <b>abdominal cancers</b><br>(Holtehdahl and colleagues, 2018) [39]                                  |  |                                       |                             |                           |   |                          |
| Holtehdahl, 2018   | Apparent performance                                   | ?                                     | ✓                           | ?                         | x   | ?                        |

**Abbreviations:** RAT (s) Risk assessment tool(s), SR2 Systematic review 2

Notes:<sup>a</sup>multiple ordered by stage of development if different

Key: ✓, low risk of bias; x, high risk of bias; ?, unclear risk of bias

risk of bias for how predictors and sample size and participants were dealt with, and a high risk of bias concerning the analysis. For the development of the other models, risk of bias was variable across all domains, although most models have a low risk of bias with respect to how outcomes are dealt with.

The external validation of the colorectal cancer RAT, and of many of the other models by Elias was judged to be of uncertain risk of bias for how sample size and patient flow was dealt with, and how analyses were conducted.

Overall, apart from the Qcancer studies, the risk of bias of the development and validation studies is mixed and/or uncertain.

### Performance of the models

As with many systematic reviews of prediction models, we found a mix of outcomes reported on the different models. The most widely reported outcome was the area under the curve (AUC). AUC estimates were calculated from external datasets for seven of the 13 models (Table 4). As some authors reported AUCs based on the model derivation dataset, in Table 4 we distinguishing between whether the reported AUC is estimated using the derivation dataset, or the external dataset. Note that for the remaining six models, which includes three of the RATs, we could find no external validation of any kind.

**Table 4** Available AUC estimates (and 95% confidence intervals) for the prediction models

| Prediction model                        | Validation (using derivation or external dataset) | Dataset used, country             | AUC (95% CI)      | Source |
|---|---|-----------------------------------|-------------------|--------|
| Colorectal cancer                       |   |                                   |                   |        |
| Bristol-Birmingham equation [29]        | Derivation  | THIN, UK                          | 0.83 (0.82, 0.84) | [29]   |
|   | External  | CAPER, UK                         | 0.92 (0.91, 0.94) | [29]   |
|   | External  | CEDAR, Netherlands                | 0.84 (0.77, 0.90) | [27]   |
| Netherlands model [28]                  | Derivation  | Primary care, Netherlands         | 0.97              | [28]   |
|   | External  | Secondary care, UK                | 0.78 (0.74, 0.81) | [40]   |
|   | External  | CEDAR, Netherlands                | 0.72 (0.62, 0.81) | [27]   |
| Netherlands model including polyps [28] | Derivation  | Primary care, Netherlands         | 0.92              | [28]   |
| Qcancer (male) [22]                     | Derivation  | Qresearch, UK                     | 0.91 (0.09, 0.91) | [22]   |
|   | External  | THIN (multiple imputation), UK    | 0.92 (0.91, 0.92) | [23]   |
|   |   | THIN (complete case analysis), UK | 0.90 (0.89, 0.91) | [23]   |
| Qcancer (female) [22]                   | Derivation  | Qresearch, UK                     | 0.89 (0.88, 0.90) | [22]   |
|   | External  | THIN (complete case analysis), UK | 0.91 (0.90, 0.92) | [23]   |
| Danish model [31]                       | External  | CEDAR, Netherlands                | 0.6 (0.48, 0.72)  | [27]   |
| RAT (2005) [33]                         | External  | CEDAR, Netherlands                | 0.81 (0.75, 0.88) | [27]   |
| Multiple cancer sites                   |   |                                   |                   |        |
| Muris abdominal complaints model [30]   | External  | CEDAR, Netherlands                | 0.62 (0.54, 0.70) | [27]   |

The Qcancer models are associated with the highest estimated AUC value from external validation: 0.92 (0.91, 0.92) and 0.91 (0.90, 0.92) for the male and female versions of the colorectal Qcancer model.

The Bristol-Birmingham equation was also associated with a high AUC value for external validity, but only in one of two studies. The two AUCs from external validation of the Bristol-Birmingham equation differ, with the AUC estimate from the UK CAPER dataset being much higher (0.92 (0.91, 0.94)) than that from the external validation using the Dutch CEDAR dataset (0.84 (0.77, 0.90)) or the derivation dataset (0.83 (0.82, 0.84)) [27, 29].

The Netherlands model for colorectal cancer was associated with the highest AUC score for internal validation (0.97), but this was not replicated when the model was used in a different population. The AUC value was much lower in both external validation studies, using either secondary care data from the UK (0.78 (0.74, 0.81)) or Dutch dataset (0.72 (0.62, 0.81)).

The remaining models are estimated to have mean AUCs between 0.6 and 0.8, with the Danish model for colorectal cancer and the Muris abdominal complaints model being the two lowest performing models. The only RAT for which an AUC is reported is for the 2005 version of the colorectal model from Elias [27], and is much lower than those from the Qcancer models, 0.81 (0.75, 0.88).

Estimates of NPV, PPV, sensitivity and specificity are available from the external validations by Elias of the Bristol-Birmingham equation [29], the models by Fijten [28], Nørrelund [31], and Muris [30] and the 2005 colorectal RAT [33]. Collins and Altman [23] also report these estimates for validation of the colorectal Qcancer model (see supplementary Table S2). The (male and female) colorectal Qcancer models are the only models to have estimates of sensitivity > 0.9 and specificity > 0.7. The 2005 colorectal RAT has a reported sensitivity of 0.95 and specificity of 0.45. The other four models (Bristol-Birmingham, Fijten, Nørrelund and Muris) all have high sensitivity (> 0.95), but very low specificity: 0.06 for Nørrelund to 0.36 for the Bristol-Birmingham equation. Marshall [29], Høltedahl [39], Hamilton [33], Hamilton [43] and Stapley [35] also report likelihood ratios (LRs), see Supplementary Table S2c. Marshall [29] report a LR of 14.7 for the Bristol-Birmingham equation, while the other 4 studies only report LRs for individual symptoms included in the model. These range from < 2 for some symptoms in the model reported in Hamilton [43] to > 30 for rectal bleeding in the model reported by Stapley [35].

#### Impact studies

Three studies were identified that attempted to evaluate the impact of tools based on diagnostic prediction models used in practice: a cross-sectional survey [16], a

pre-post study [55] and a randomised controlled trial [56]. The RCT and pre-post studies evaluated the use of a combination of tools which included RATs for colorectal cancer. The cross-sectional survey by Price [16] evaluated the impact of GP practice access to RAT and/or Qcancer, see Table 5.

Price and colleagues [16] compared UK practice-level 2WW referral rates between GP practices that reported access to RAT and/or Qcancer, with practices that reported no access to these two tools. The tools included Qcancer and RAT for any cancer, and the analyses were not restricted to colorectal cancer.

Hamilton and colleagues (2013) [55] investigated the number of times two RATs [34] – one for lung and one for colorectal cancer – were used, together with the number of subsequent referrals and investigations, before and 6 months after the introduction of the tools in general practice in the UK.

Emery and colleagues (2017) [56] evaluated the impact of two complex interventions in rural Australia – a GP intervention and a cancer awareness campaign – in a 2 × 2 design trial, compared to control groups. The GP intervention consisted of an “education resource card” that included RATs for colorectal, lung and prostate cancer, together with summaries of relevant guidelines for colorectal, lung and prostate cancer, with the addition of guidelines for breast cancer and training on the use of these resources. The RATs were based on diagnostic prediction models developed using a patient cohort from the UK [34]. Emery and colleagues (2017) [56] used the total diagnostic interval (TDI), i.e. the time from first symptom to cancer diagnosis, as an outcome measure.

**Critical appraisal**

The RCT by Emery was found to be at low risk of bias (see Table 6). Given the observational nature of the studies by Hamilton and Price [16], there are a number of concerns regarding their risk of bias.

**Study outcomes**

Emery and colleagues [56] did not find significant differences in the median or log-transformed (ln) mean time to diagnosis at either intervention level (community intervention vs control, GP intervention vs control) or when analysed by factorial design, tumour group or sub-intervals of the TDI.

Hamilton and colleagues (2013) [55] reported on changes in investigations carried out and rapid referrals before and after the introduction of the tools. They found a 26% increase in referrals for colorectal cancer and a 15% increase in GP requests for colonoscopies after introduction of the tools. However, only absolute numbers are reported, without data on total numbers of patients and GP visits, or the appropriateness of the referral.

Price and colleagues [16] did not find any differences in mean 2WW referral rates between practices reporting access to cancer decision-making tools and those who did not: mean difference in referral rate of 3.1 per 100, 000 population (95% CI of – 5.5, 11.7). As the study considered RATs and Qcancer for any suspected cancer and 2WW referral rates for any cancer, the specific impact of colorectal cancer-relevant RATs or Qcancer tools on referrals for colorectal cancer cannot be evaluated.

Study results are summarised in Table 7.

**Discussion**

This review summarised existing evidence on development, validation, accuracy and impact of prediction models developed to help diagnosis of colorectal cancer in primary care. A large number of prediction models were identified consisting of one-off models and models from the RAT and Qcancer series. Validation and impact assessment of these models in appropriate settings is currently limited, and we found no economic evaluations of any tools.

Currently, most research on developing symptom-based colorectal cancer risk prediction models is concentrated in

**Table 5** Description of tools assessed in the three impact studies

| Study ID                           | Prediction tool   | Country of tool development      | Tool description   |
|------------------------------------|---|----------------------------------|--|
| Hamilton and colleagues, 2013 [55] | RAT presented on a mouse mat and desk top flip chart (for lung and colorectal cancer) | UK                               | The RAT algorithm is displayed in a table/matrix format, which allows a risk estimate to be calculated for a single symptom, pairs of symptoms or repeat attendances with the same symptom. The values are colour-coded to aid interpretation. |
| Emery and colleagues, 2017 [56]    | Education resource card containing the RAT and referral guidelines                    | UK (RAT), Australia (guidelines) | Resource card containing the RAT tables for colorectal, lung and prostate cancer, as well as the Australian National Breast and Ovarian Cancer Centre guidelines for investigating new breast symptoms   |
| Price and colleagues 2019 [16]     | RAT and/or Qcancer in any form (e.g. paper, software etc.) for any cancer             | UK                               | Any affirmative GP practice access to RAT and/or Qcancer   |

Abbreviations: ID Identification, RAT(s) Risk assessment tool(s)

**Table 6** Risk of bias assessment for the three impact studies

|                                     | Random sequence generation | Allocation concealment | Baseline outcome measurements similar | Baseline characteristics similar | Incomplete outcome data | Knowledge of the allocated interventions adequately prevented during the study | Protection against contamination | Selective outcome reporting | Other risks of bias |
|-------------------------------------|----------------------------|------------------------|---------------------------------------|----------------------------------|-------------------------|--|----------------------------------|-----------------------------|---------------------|
| <b>Randomised controlled trials</b> |                            |                        |                                       |                                  |                         |  |                                  |                             |                     |
| Emery 2017 [56]                     | ✓                          | x                      | ✓                                     | ✓                                | ✓                       | ✓  | ✓                                | ✓                           | ✓                   |
| <b>Pre-post study</b>               |                            |                        |                                       |                                  |                         |  |                                  |                             |                     |
| Hamilton 2013 [55]                  | N/A                        | N/A                    | N/A                                   | N/A                              | ?                       | N/A  | N/A                              | ?                           | x                   |
| <b>Cross-sectional survey</b>       |                            |                        |                                       |                                  |                         |  |                                  |                             |                     |
| Price and colleagues 2019 [16]      | N/A                        | N/A                    | N/A                                   | ?                                | ?                       | ✓  | N/A                              | ✓                           | x                   |

Abbreviations: N/A Not applicable. Key: ✓, low risk; x, high risk; ?, unclear risk

Europe and, in particular, the UK. Qcancer and RAT are the dominant prediction models, and highlight important knowledge gaps: the Qcancer models are developed on higher quality data (cohort data) than the RATs, and have been externally validated, but lack specific impact assessment. In contrast, the RAT models have more evidence of impact in practice, but were developed from case-control studies and have limited external validation. Ideally, this is an area for further development of the RATs, and the other models that had not been externally validated. This lack of evaluation seems consistent with prediction models in other disease areas [57].

Other systematic reviews have looked at feature-based cancer diagnostic tools in primary care. Williams and

colleagues (2016) [17] conducted a systematic review of studies that described, validated or assessed the impact of colorectal cancer diagnostic tools. They identified reports on the development and/or validation of 15 models: nine relevant to primary care and six for secondary care. They also identified one study looking at referral patterns (for colorectal cancer RAT [55]). However, they did not identify any studies that tested whether patients who were diagnosed with the aid of the tool fared better than those who were diagnosed without it. In a similar review, looking at risk prediction models for screening, Usher-Smith and colleagues (2015) [53] concluded that, even though some of the colorectal cancer prediction models had potential for clinical application, there remains considerable

**Table 7** Results reported by the impact studies

| Study ID           | Prediction tool   | Country   | Study design                                | Intended purpose  | Main results for colorectal RAT   |
|--------------------|---|-----------|---|---|---|
| Hamilton 2013 [55] | RAT for lung, colorectal cancer in two formats: mouse mat and desk top flip chart | UK        | Pre-post study                              | To compare referrals and investigations for colorectal and lung cancer before and after the implementation of RATs  | 26% increase in 2-week referrals (1173 before, 1477 after); 15% increase in colonoscopies (1762 before, 2032 after)<br><b>No conclusion possible on the effectiveness of the intervention</b>   |
| Emery 2017 [56]    | Education resource card including RAT for colorectal, lung and prostate cancer    | Australia | Factorial cluster RCT                       | to measure the effect of community-based symptom awareness and GP-based educational interventions on the time to diagnosis (i.e. TDI) for patients presenting with breast, prostate, colorectal or lung cancer in rural Western Australia | <b>No significant differences in the median or in mean TDI at either intervention level:</b><br><b>Colorectal cancer:</b><br>-GP intervention vs control: median TDI 124 vs 122 days; In mean difference - 0.03 95% CI - 0.51-0.45 P = 0.42<br>-community intervention vs control: median TDI 107 vs 133 days; In mean difference - 0.26 95% CI - 0.63-0.11 P = 0.16; |
| Price 2019 [16]    | Access to any RAT and/or Qcancer tool in any format                               | UK        | Cross-sectional survey at GP practice level | To compare the mean 2WW referral rates between GP practices reporting access to RAT and/or Qcancer and those who reported no access to these tools  | <b>No statistically significant difference between mean referral rates between practices reporting access or no access to RAT and/or Qcancer:</b> mean difference of 3.1 referrals per 100,000 population (95% CI - 5.5, 11.7, p-value 0.48)  |

Abbreviations: ANOVA Analysis of variance, GP General practitioner, NHS National Health Service, OR Odds ratio, RAT(s) Risk assessment tool(s), RCT Randomised controlled trial, SRT Systematic review 1, TDI Total diagnostic interval, UK United Kingdom

uncertainty about their clinical utility. Similarly, Schmidt-Hansen and colleagues (2017) [58] conducted a review of lung cancer tools and found limited evidence to support the recommendation of any of the identified risk prediction tools, due to lack of external validation or cost impact assessment.

Our systematic review identified two impact studies, published after the review by Williams et al. [17], both of which indicating little evidence of an impact from using these tools in primary care. However, it is still difficult to conclude whether these tools have any impact on patient outcomes. For instance, concerns on the quality of the studies makes it unclear whether the lack of effect was due to poor implementation of the tools in practice, insufficient uptake by the GPs or limited marginal contribution of the tools in assessing the risk of cancer. The best quality study (Emery and colleagues 2017 [56]) failed to show a significant effect; however, the composite intervention used, combining older versions of several instruments (developed on populations from a different country), could have limited the effectiveness of the diagnostic tools. Thus, there is still a need for good quality studies to examine the impact of using prediction model based tools to help colorectal cancer diagnosis in primary care.

Only prediction models were included in our systematic review. Other aids, such as algorithms or guidelines may be useful, but were excluded from this review. However, the systematic review by Elias et al. [27] had a much broader inclusion criteria for “model”. The review found a previous version of the NICE guidelines to be the best performing (when validated against the CEDAR dataset). Importantly, this review did not include any of the Qcancer models, which are associated with AUCs greater than those reported for the NICE guidelines.

The systematic review followed a pre-specified protocol, and the team conducting the review are independent and experienced in systematic review methodology.

Our findings are limited by the quality of the studies included in the systematic review, in particular, among the limitations of the impact studies were lack of randomisation, lack of patient-related outcomes and use of tools on populations they were not developed for (e.g. use of a UK-developed tool on an Australian population). The outcome measures used by some of the impact studies make it difficult to interpret reports of an increase in referral rate without including reasonable assessment of the appropriateness of the referral or subsequent impact on cancer vs non-cancer diagnosis.

## Conclusion

Current evaluations provide limited evidence of the impact on patient outcomes of using feature-based cancer diagnostic tools in primary care. The lack of robust

effectiveness data is also likely to be a major limiting factor in assessing their cost-effectiveness. More research is needed to externally validate prediction models that could be used as tools, as well as more research on the impact of using these tools in clinical practice. However, choice of study design and outcomes for future evaluations of the impact of tools, may not be straightforward. Practical reasons may highlight the potential need for a cluster and pragmatic trial design. Arguably, by comparing average times to diagnosis, patients not prioritised for quick referrals are less at risk of being missed. The debate, however, is ongoing on the most appropriate outcomes for evaluating interventions to improve cancer diagnosis and referral.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12885-020-07572-z>.

**Additional file 1: Table S1.** MEDLINE literature search strategy. **Table S2a.** Development and validation studies - Characteristics (1). **Table S2b.** Development and validation studies - Characteristics (2). **Table S2c.** Development and validation studies - Model development and performance. **Table S2d.** Development and validation studies - Results. **Table S3a.** Development and validation studies - Risk of bias assessment, Questions 1 to 3. **Table S3b.** Development and validation studies - Risk of bias assessment, Questions 4 to 5. **Table S4a.** Impact Studies - Characteristics. **Table S4b.** Impact studies - Study Design. **Table S4c.** Impact studies - Results. **Table S5.** Impact studies - Critical Appraisal.

## Abbreviations

AUC: Area under the curve; CAPER: Cancer Prediction in Exeter; CEDAR: Cost-Effectiveness of a Decision rule for Abdominal complaints in Primary care; CHARMS: CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies; CRC: Colorectal cancer; eCDS: Electronic Cancer Decision Support; EMIS: Egton Medical Information Systems; EPOC: Effective Practice and Organisation of Care; GP(s): General practitioner(s); NICE: National Institute for Health and Care Excellence; PRIS MA: Reporting Items for Systematic Reviews and Meta-Analyses; RAT(s): Risk Assessment Tool(s); RCT: Randomised controlled trial; TDI: Total diagnostic interval; THIN: The Health Improvement Network; UK: United Kingdom; WW: Week wait

## Acknowledgements

The authors would like to thank Ms. Jenny Lowe for her support and Dr. Chris Cooper for his contribution.

## Registration

PROSPERO: Assessing the impact of diagnostic prediction tools for cancer in primary care: a systematic review is registered as PROSPERO CRD42017068373 and Prediction models for aiding cancer diagnosis in primary care: a systematic review is registered as PROSPERO CRD42017068375.

## Authors' contributions

BG contributed to the conception, design, interpretation of data and drafted the first version of the manuscript. RL participated in data extraction and analysis, contributed to the conception, design and interpretation of data. JP participated in data extraction and analysis, contributed to the conception, design and interpretation of data. SR carried out the electronic searches for the review, contributed to the conception, design and interpretation of data. CJH is the guarantor of the study, he developed the concept and protocol, and assisted with interpretation of data. All authors agreed the final version of paper.

**Funding**

This report was commissioned by the NIHR HTA Programme as project number 16/12/04.

**Availability of data and materials**

The data that support the findings of this study are available within the article or its supplementary materials.

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

All authors have no conflicts of interest to disclose.

**Author details**

<sup>1</sup>Exeter Test Group, College of Medicine and Health, University of Exeter Medical School, Exeter, UK. <sup>2</sup>North Wales Centre for Primary Care Research, Bangor University, Bangor, UK. <sup>3</sup>Peninsula Technology Assessment Group, College of Medicine and Health, University of Exeter Medical School, Exeter, UK.

Received: 5 February 2020 Accepted: 26 October 2020

Published online: 10 November 2020

**References**

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68:394–424.
- Bowel cancer statistics [<https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer>].
- Ades AE, Biswas M, Welton NJ, Hamilton W. Symptom lead time distribution in lung cancer: natural history and prospects for early diagnosis. *Int J Epidemiol*. 2014;43(6):1865–73.
- Cole SR, Tucker GR, Osborne JM, Byrne SE, Bampton PA, Fraser RJ, Young GP. Shift to earlier stage at diagnosis as a consequence of the National Bowel Cancer Screening Program. *Med J Aust*. 2013;198(6):327–30.
- Richards MA, Westcombe AM, Love SB, Littlejohns P, Ramirez AJ. Influence of delay on survival in patients with breast cancer: a systematic review. *Lancet*. 1999;353(9159):1119–26.
- Elliss-Brookes L, McPhail S, Ives A, Greenslade M, Shelton J, Hiom S, Richards M. Routes to diagnosis for cancer—determining the patient journey using multiple routine data sets. *Br J Cancer*. 2012;107(8):1220–6.
- Richards M. The national awareness and early diagnosis initiative in England: assembling the evidence. *Br J Cancer*. 2009;101(Suppl 2):S1.
- National Institute for Health Care. Suspected cancer: recognition and referral. In: NICE guidelines NG12; 2015.
- Jones CP, Fallaize RC, Longman RJ. Updated 'two-week wait' referral guidelines for suspected colorectal cancer have increased referral volumes without improving cancer detection rates. *Br J Med Pract*. 2019;12(2):a012.
- Vulliamy P, McCluney S, Raouf S, Banerjee S. Trends in urgent referrals for suspected colorectal cancer: an increase in quantity, but not in quality. *Ann R Coll Surg Engl*. 2016;98(8):564–7.
- Hiom S. Diagnosing cancer earlier: reviewing the evidence for improving cancer survival. *Br J Cancer*. 2015;112:S1–5.
- Lyratzopoulos G, Neal RD, Barbiere JM, Rubin GP, Abel GA. Variation in number of general practitioner consultations before hospital referral for cancer: findings from the 2010 National Cancer Patient Experience Survey in England. *Lancet Oncol*. 2012;13(4):353–65.
- Hendriksen JM, Geersing G-J, Moons KG, de Groot JA. Diagnostic and prognostic prediction models. *J Thromb Haemost*. 2013;11:129–41.
- Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162(1):W1–W73.
- Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, Riley RD, Hemingway H, Altman DG, Group P. Prognosis research strategy (PROGRESS) 3: prognostic model research. *PLoS Med*. 2013;10(2):e1001381.
- Price S, Spencer A, Medina-Lara A, Hamilton W. Availability and use of cancer decision-support tools: a cross-sectional survey of UK primary care. *Br J Gen Pract*. 2019;69(684):e437–43.
- Williams TGS, Cubiella J, Griffin SJ, Walter FM, Usher-Smith JA. Risk prediction models for colorectal cancer in people with symptoms: A systematic review. *BMC Gastroenterol*. 2016;16(1).
- Medina-Lara A, Grigore B, Lewis R, Peters J, Price S, Landa P, Robinson S, Neal R, Hamilton W, Spencer A. Understanding the effectiveness, cost-effectiveness and current use of cancer diagnostic tools to aid decision-making in primary care. In: *Health Technology Assessment: National Institute for Health Research*; 2020. <https://www.journalslibrary.nihr.ac.uk/programmes/hta/161204/#/>.
- Centre for R. Dissemination: Systematic Reviews: CRD's guidance for undertaking reviews in health care. York: Centre for Reviews and Dissemination; 2009.
- Moher D, Liberati A, Tetzlaff J, Altman DG. The PRISMA statement: Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097.
- Bouwmeester W, Zuithoff NPA, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, Altman DG, Moons KGM. Reporting and methods in clinical prediction research: a systematic review. *PLoS Med*. 2012;9(5):e1001221.
- Hippisley-Cox J, Coupland C. Identifying patients with suspected colorectal cancer in primary care: derivation and validation of an algorithm. *Br J Gen Pract*. 2012;62(594):e29–37.
- Collins GS, Altman DG. Identifying patients with undetected colorectal cancer: an independent validation of Qcancer (colorectal). *Br J Cancer*. 2012;107(2):260–5.
- Moons KGM, de Groot JAH, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, Reitsma JB, Collins GS. Critical appraisal and data extraction for systematic reviews of prediction Modelling studies: the CHARMS checklist. *PLoS Med*. 2014;11(10):e1001744.
- Wolff R, Whiting P, Mallett S. PROBAST: a risk of bias tool for prediction modelling studies. In: *Cochrane Colloquium Vienna*, vol. 2015; 2015.
- Higgins JPT, Altman DG, Sterne JAC: Chapter 8: Assessing risk of bias in included studies. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1*. The Cochrane Collaboration 2011. 0 [updated March 2011].
- Elias SG, Kok L, Witteman BJ, Goedhard JG, Romberg-Camps MJ, Muris JW, de Wit NJ, Moons KG. Published diagnostic models safely excluded colorectal cancer in an independent primary care validation study. *J Clin Epidemiol*. 2017;82:149–57 e148.
- Fijten GH, Starmans R, Muris JW, Schouten HJ, Blijham GH, Knottnerus JA. Predictive value of signs and symptoms for colorectal cancer in patients with rectal bleeding in general practice. *Fam Pract*. 1995;12(3):279–86.
- Marshall T, Lancashire R, Sharp D, Peters TJ, Cheng KK, Hamilton W. The diagnostic performance of scoring systems to identify symptomatic colorectal cancer compared to current referral guidance. *Gen*. 2011;60(9):1242–8.
- Muris JW, Starmans R, Fijten GH, Crebolder HF, Schouten HJ, Knottnerus JA. Non-acute abdominal complaints in general practice: diagnostic value of signs and symptoms. *Br J Gen Pract*. 1995;45(395):313–6.
- Nørrelund N, Nørrelund H. Colorectal cancer and polyps in patients aged 40 years and over who consult a GP with rectal bleeding. *Fam Pract*. 1996; 13(2):160–5.
- Kop R, Hoogendoorn M, Teije AT, Büchner FL, Slottje P, Moons LMG, Numans ME. Predictive modeling of colorectal cancer using a dedicated pre-processing pipeline on routine electronic medical records. *Comput Biol Med*. 2016;76:30–8.
- Hamilton W, Round A, Sharp D, Peters TJ. Clinical features of colorectal cancer before diagnosis: a population-based case-control study. *Br J Cancer*. 2005;93(4):399–405.
- Hamilton W. The CAPER studies: five case-control studies aimed at identifying and quantifying the risk of cancer in symptomatic primary care patients. *Br J Cancer*. 2009;101:S80–6.
- Stapley SA, Rubin GP, Alsina D, Shephard EA, Rutter MD, Hamilton WT. Clinical features of bowel disease in patients aged <50 years in primary care: a large case-control study. *Br J Gen Pract*. 2017;67(658):e336–44.
- Hamilton W, Barrett J, Stapley S, Sharp D, Rose P. Clinical features of metastatic cancer in primary care: a case-control study using medical records. *Br J Gen Pract*. 2015;65(637):e516–22.
- 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(606):e1–e10.

38. 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(606):e11–21.
39. Holtehdahl K, Hjertholm P, Borgquist L, Donker GA, Buntinx F, Weller D, Braaten T, Månsson J, Strandberg EL, Campbell C. Abdominal symptoms and cancer in the abdomen: prospective cohort study in European primary care. *Br J Gen Pract.* 2018;68(670):e301–10.
40. Hodder RJ, Ballal M, Selvachandran S, Cade D. Pitfalls in the construction of cancer guidelines demonstrated by the analyses of colorectal referrals. *Ann R Coll Surg Engl.* 2005;87(6):419–26.
41. Kop R, Hoogendoorn M, Moons LMG, Numans ME, ten Teije A. On the advantage of using dedicated data mining techniques to predict colorectal cancer. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9105; 2015. p. 133–42.
42. Hoogendoorn M, Szolovits P, Moons LMG, Numans ME. Utilizing uncoded consultation notes from electronic medical records for predictive modeling of colorectal cancer. *Artif Intell Med.* 2015;69:53–61.
43. Hamilton W, Lancashire R, Sharp D, Peters TJ, Cheng K, Marshall T. The risk of colorectal cancer with symptoms at different ages and between the sexes: a case-control study. *BMC Med.* 2009;7:17.
44. Stapley S, Peters TJ, Neal RD, Rose PW, Walter FM, Hamilton W. The risk of oesophago-gastric cancer in symptomatic patients in primary care: a large case-control study using electronic records. *Br J Cancer.* 2013;108(1):25–31.
45. Hamilton W. Cancer diagnosis in primary care. *Br J Gen Pract.* 2010;60(571):121–8.
46. Shephard E, Neal R, Rose P, Walter F, Hamilton WT. Clinical features of kidney cancer in primary care: a case-control study using primary care records. *Br J Gen Pract.* 2013;63(609):e250–5.
47. Shephard EA, Hamilton W, Neal RD, Rose PW, Walter FM. Symptoms of adult chronic and acute leukaemia before diagnosis: large primary care case-control studies using electronic records. *Br J Gen Pract.* 2016;66(644):e182–8.
48. Shephard EA, Neal RD, Rose P, Walter FM, Litt EJ, Hamilton WT. Quantifying the risk of multiple myeloma from symptoms reported in primary care patients: a large case-control study using electronic records. *Br J Gen Pract.* 2015;65(631):e106–13.
49. Shephard EA, Stapley S, Neal RD, Rose P, Walter FM, Hamilton WT. Clinical features of bladder cancer in primary care. *Br J Gen Pract.* 2012;62(602):e598–604.
50. Stapley S, Peters TJ, Neal RD, Rose PW, Walter FM, Hamilton W. The risk of pancreatic cancer in symptomatic patients in primary care: a large case-control study using electronic records. *Br J Cancer.* 2012;106(12):1940–4.
51. Walker S, Hyde C, Hamilton W. Risk of uterine cancer in symptomatic women in primary care: case-control study using electronic records. *Br J Gen Pract.* 2013;63(614):e643–8.
52. Hamilton W, Lancashire R, Sharp D, Peters TJ, Cheng KK, Marshall T. The importance of anaemia in diagnosing colorectal cancer: a case-control study using electronic primary care records. *Br J Cancer.* 2008;98(2):323–7.
53. Usher-Smith J, Emery J, Hamilton W, Griffin SJ, Walter FM. Risk prediction tools for cancer in primary care. *Br J Cancer.* 2015;113:1645.
54. Lewis JD, Schinnar R, Bilker WB, Wang X, Strom BL. Validation studies of the health improvement network (THIN) database for pharmacoepidemiology research. *Pharmacoepidemiol Drug Saf.* 2007;16(4):393–401.
55. Hamilton W, Green T, Martins T, Elliott K, Rubin G, Macleod U. Evaluation of risk assessment tools for suspected cancer in general practice: a cohort study. *Br J Gen Pract.* 2013;63(606):e30–6.
56. Emery JD, Gray V, Walter FM, Cheetham S, Croager EJ, Slevin T, Saunders C, Threlfall T, Auret K, Nowak AK, et al. The improving rural Cancer outcomes trial: a cluster-randomised controlled trial of a complex intervention to reduce time to diagnosis in rural cancer patients in Western Australia. *Br J Cancer.* 2017;117(10):1459–69.
57. van Giessen A, Peters J, Wilcher B, Hyde C, Moons C, de Wit A, Koffijberg E. Systematic review of health economic impact evaluations of risk prediction models: stop developing, Start Evaluating. *Value Health.* 2017;20(4):718–26.
58. Schmidt-Hansen M, Berendse S, Hamilton W, Baldwin DR. Lung cancer in symptomatic patients presenting in primary care: a systematic review of risk prediction tools. *Br J Gen Pract.* 2017;67(659):e396–s404.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

