NIHR | National Institute for Health Research

Journals Library

Check for updates

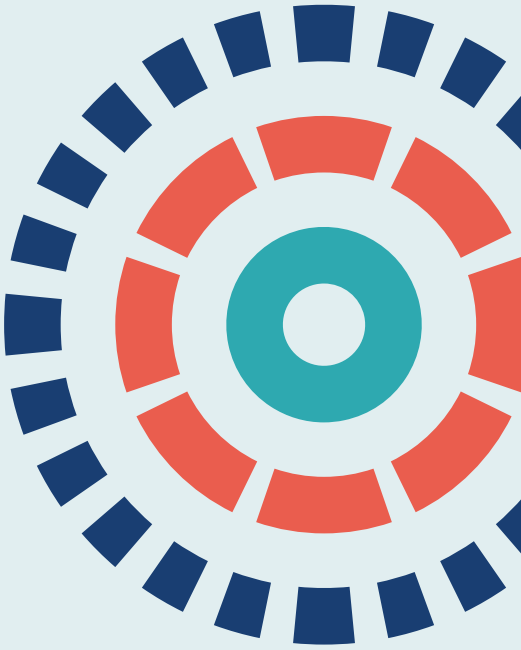
Health Technology Assessment

Volume 24 • Issue 66 • November 2020

ISSN 1366-5278

Cancer diagnostic tools to aid decision-making in primary care: mixed-methods systematic reviews and cost-effectiveness analysis

Antonieta Medina-Lara, Bogdan Grigore, Ruth Lewis, Jaime Peters, Sarah Price, Paolo Landa, Sophie Robinson, Richard Neal, William Hamilton and Anne E Spencer



DOI 10.3310/hta24660

Cancer diagnostic tools to aid decision-making in primary care: mixed-methods systematic reviews and cost-effectiveness analysis

Antonieta Medina-Lara,[1*] Bogdan Grigore,[2] Ruth Lewis,[3] Jaime Peters,[2] Sarah Price,[4] Paolo Landa,[1] Sophie Robinson,[5] Richard Neal,[6] William Hamilton[4] and Anne E Spencer[1]

[1]Health Economics Group, College of Medicine and Health, University of Exeter Medical School, Exeter, UK
[2]Exeter Test Group, College of Medicine and Health, University of Exeter Medical School, Exeter, UK
[3]North Wales Centre for Primary Care Research, Bangor University, Bangor, UK
[4]Primary Care Diagnostics, College of Medicine and Health, University of Exeter Medical School, Exeter, UK
[5]Peninsula Technology Assessment Group, College of Medicine and Health, University of Exeter Medical School, Exeter, UK
[6]Leeds Institute of Health Sciences, University of Leeds, Leeds, UK

*Corresponding author

Declared competing interests of authors: William Hamilton has overseen the development of a suite of cancer risk assessment tools encompassing all the major adult cancers. Richard Neal has also contributed to some of these studies. The risk assessment tools are available at no cost to the NHS. William Hamilton is the chief investigator of the Electronic Risk Assessment Tools for Cancer (ERICA) trial, a philanthropically funded cluster randomised controlled trial of electronic risk assessment tools in primary care. As a result of this interest, William Hamilton excluded himself from the data analysis, although he contributed to the rest of the work, including writing the outputs. Anne E Spencer and Antonieta Medina-Lara also report supporting the ERICA trial. William Hamilton, Antonieta Medina-Lara and Anne E Spencer report grants from Gillings Foundation and minor support from Cancer Research UK for the ERICA trial. Antonieta Medina-Lara reports grants from the National Institute for Health Research during the conduct of the study and outside the submitted work.

Published November 2020
DOI: 10.3310/hta24660

This report should be referenced as follows:

Medina-Lara A, Grigore B, Lewis R, Peters J, Price S, Landa P, *et al*. Cancer diagnostic tools to aid decision-making in primary care: mixed-methods systematic reviews and cost-effectiveness analysis. *Health Technol Assess* 2020;**24**(66).

Health Technology Assessment is indexed and abstracted in *Index Medicus*/MEDLINE, *Excerpta Medica*/EMBASE, *Science Citation Index Expanded* (SciSearch®) and *Current Contents®*/ Clinical Medicine.

Health Technology Assessment

HTA/HTA TAR

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 3.370

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, the Cochrane Library and Clarivate Analytics Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

Health Technology Assessment (HTA) research is undertaken where some evidence already exists to show that a technology can be effective and this needs to be compared to the current standard intervention to see which works best. Research can evaluate any intervention used in the treatment, prevention or diagnosis of disease, provided the study outcomes lead to findings that have the potential to be of direct benefit to NHS patients. Technologies in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 16/12/04. The contractual start date was in April 2017. The draft report began editorial review in March 2019 and was accepted for publication in March 2020. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded by the National Institute for Health Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, NETSCC, the HTA programme or the Department of Health and Social Care.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

Abstract

Cancer diagnostic tools to aid decision-making in primary care: mixed-methods systematic reviews and cost-effectiveness analysis

Antonieta Medina-Lara,[1*] Bogdan Grigore,[2] Ruth Lewis,[3] Jaime Peters,[2] Sarah Price,[4] Paolo Landa,[1] Sophie Robinson,[5] Richard Neal,[6] William Hamilton[4] and Anne E Spencer[1]

[1]Health Economics Group, College of Medicine and Health, University of Exeter Medical School, Exeter, UK
[2]Exeter Test Group, College of Medicine and Health, University of Exeter Medical School, Exeter, UK
[3]North Wales Centre for Primary Care Research, Bangor University, Bangor, UK
[4]Primary Care Diagnostics, College of Medicine and Health, University of Exeter Medical School, Exeter, UK
[5]Peninsula Technology Assessment Group, College of Medicine and Health, University of Exeter Medical School, Exeter, UK
[6]Leeds Institute of Health Sciences, University of Leeds, Leeds, UK

*Corresponding author A.Medina-Lara@exeter.ac.uk

Background: Tools based on diagnostic prediction models are available to help general practitioners diagnose cancer. It is unclear whether or not tools expedite diagnosis or affect patient quality of life and/or survival.

Objectives: The objectives were to evaluate the evidence on the validation, clinical effectiveness, cost-effectiveness, and availability and use of cancer diagnostic tools in primary care.

Methods: Two systematic reviews were conducted to examine the clinical effectiveness (review 1) and the development, validation and accuracy (review 2) of diagnostic prediction models for aiding general practitioners in cancer diagnosis. Bibliographic searches were conducted on MEDLINE, MEDLINE In-Process, EMBASE, Cochrane Library and Web of Science) in May 2017, with updated searches conducted in November 2018. A decision-analytic model explored the tools' clinical effectiveness and cost-effectiveness in colorectal cancer. The model compared patient outcomes and costs between strategies that included the use of the tools and those that did not, using the NHS perspective. We surveyed 4600 general practitioners in randomly selected UK practices to determine the proportions of general practices and general practitioners with access to, and using, cancer decision support tools. Association between access to these tools and practice-level cancer diagnostic indicators was explored.

Results: Systematic review 1 – five studies, of different design and quality, reporting on three diagnostic tools, were included. We found no evidence that using the tools was associated with better outcomes. Systematic review 2 – 43 studies were included, reporting on prediction models, in various stages of development, for 14 cancer sites (including multiple cancers). Most studies relate to QCancer® (ClinRisk Ltd, Leeds, UK) and risk assessment tools.

Decision model: In the absence of studies reporting their clinical outcomes, QCancer and risk assessment tools were evaluated against faecal immunochemical testing. A linked data approach was used, which translates diagnostic accuracy into time to diagnosis and treatment, and stage at

diagnosis. Given the current lack of evidence, the model showed that the cost-effectiveness of diagnostic tools in colorectal cancer relies on demonstrating patient survival benefits. Sensitivity of faecal immunochemical testing and specificity of QCancer and risk assessment tools in a low-risk population were the key uncertain parameters.

Survey: Practitioner- and practice-level response rates were 10.3% (476/4600) and 23.3% (227/975), respectively. Cancer decision support tools were available in 83 out of 227 practices (36.6%, 95% confidence interval 30.3% to 43.1%), and were likely to be used in 38 out of 227 practices (16.7%, 95% confidence interval 12.1% to 22.2%). The mean 2-week-wait referral rate did not differ between practices that do and practices that do not have access to QCancer or risk assessment tools (mean difference of 1.8 referrals per 100,000 referrals, 95% confidence interval –6.7 to 10.3 referrals per 100,000 referrals).

Limitations: There is little good-quality evidence on the clinical effectiveness and cost-effectiveness of diagnostic tools. Many diagnostic prediction models are limited by a lack of external validation. There are limited data on current UK practice and clinical outcomes of diagnostic strategies, and there is no evidence on the quality-of-life outcomes of diagnostic results. The survey was limited by low response rates.

Conclusion: The evidence base on the tools is limited. Research on how general practitioners interact with the tools may help to identify barriers to implementation and uptake, and the potential for clinical effectiveness.

Future work: Continued model validation is recommended, especially for risk assessment tools. Assessment of the tools' impact on time to diagnosis and treatment, stage at diagnosis, and health outcomes is also recommended, as is further work to understand how tools are used in general practitioner consultations.

Study registration: This study is registered as PROSPERO CRD42017068373 and CRD42017068375.

Funding: This project was funded by the National Institute for Health Research (NIHR) Health Technology programme and will be published in full in *Health Technology Assessment*; Vol. 24, No. 66. See the NIHR Journals Library website for further project information.

Contents

List of tables xiii

List of figures xvii

List of boxes xix

List of abbreviations xxi

Plain English summary xxiii

Scientific summary xxv

Chapter 1 Introduction 1
Scientific background and rationale 1
Aims and objectives 3

Chapter 2 Systematic reviews 1 and 2: literature search strategy 5
Introduction 5
Methods 5
Search strategy 6
Inclusion and exclusion criteria 7
Selection of studies 8
Results 8
Studies identified 8

Chapter 3 Systematic review 1 11
Objective 11
Methods 11
Identification of studies 11
Data extraction 11
Critical appraisal 11
Data synthesis 11
Results 11
Studies identified 11
Critical appraisal 13
Study outcomes 13
Discussion 13
Conclusion 19

Chapter 4 Systematic review 2 21
Objective 21
Methods 21
Identification of studies 21
Data extraction 21
Critical appraisal 21
Data synthesis 21

Results 21
Studies identified 21
Critical appraisal 29
Discussion 29
Conclusion 29

Chapter 5 Updated review **31**
Introduction 31
Aim 32
Methods 32
Identification of studies 32
Data extraction 33
Critical appraisal 33
Data synthesis 35
Results 35
Studies identified 35
Findings of the previous review 35
Findings of the updated review 35
Colorectal cancer 40
The overall findings 41
Discussion 44
Summary of the findings 44
Existing reviews of colorectal cancer 44
Strengths and limitations of the updated review 44
Conclusion 46

Chapter 6 Data for informing the economic decision model **47**
Objective 47
Sources of information 47
Summary of findings for the decision-analytic model 47
Systematic review 1 47
Systematic review 2 48
Updated review 49
Systematic review of existing economic decision-analytic models for colorectal cancer 50
Methods 50
Search strategy 50
Results 52
Conclusions 64

Chapter 7 Economic decision-analytic model for colorectal cancer **65**
Aim and objectives 65
Methods 66
Conceptualisation of the model 66
Population 70
Interventions 70
Comparators 73
Diagnostic pathway model 73
Disease history model linking test results and clinical outcomes 74
Results 83
Colorectal cancer patients 83
Whole-population analysis 83
Tornado analysis 85
Probabilistic analysis 85

Threshold analysis 86
Scenario analyses 87
Discussion 89
Conclusion 90

Chapter 8 General practitioner survey **95**
Introduction 95
Existing research 96
Research aims 96
Methods 97
Questionnaire design 97
Sample 97
Survey administration 97
Analyses 98
Results 98
Sample characteristics 98
Paper-based risk assessment tools 100
Electronic clinical decision support tools 101
Combined tools 102
Association between the use of tools and 2-week-wait referral activity 102
Discussion 103
Strengths and limitations 103
What this research adds 104
Conclusion 104

Chapter 9 Discussion **105**
What evidence exists on the clinical effectiveness and cost-effectiveness of cancer diagnostic tools in primary care? (Systematic review 1) 105
What diagnostic prediction models exist in the literature that have the potential to be used as tools in primary care? (Systematic review 2) 105
What evidence exists on the association between different diagnostic interval durations and patient outcomes? (Updated review) 105
What is the likely impact that use of the tools in clinical practice may have on patient outcomes and NHS resources? (Economic model) 106
To what extent do general practitioners have access to and use existing cancer decision support tools? (General practitioner survey) 107
Strengths 107
Limitations 107
Uncertainties 108

Chapter 10 Conclusions **111**
Recommendations for practice 111
Recommendations for research 111

Acknowledgements **113**

References **115**

Appendix 1 Literature search strategies **143**

Appendix 2 Systematic review 1: supplementary tables **147**

Appendix 3 Systematic review 2: supplementary tables **153**

Appendix 4 Updated review: supplementary tables **211**

Appendix 5 Economic decision-analytic model **283**

Appendix 6 General practitioner survey **321**

List of tables

TABLE 1 Stages of development of a diagnostic prediction model 6

TABLE 2 Search strategy for MEDLINE (searched May 2017) 7

TABLE 3 Inclusion and exclusion criteria 9

TABLE 4 Systematic review 1: description of the tools assessed 12

TABLE 5 Systematic review 1: risk-of-bias assessment 14

TABLE 6 Systematic review 1: results reported by the studies 15

TABLE 7 Risk-of-bias assessment form based on PROBAST (based on Wolff *et al.*) 22

TABLE 8 Systematic review 2: reviews included 22

TABLE 9 Systematic review 2: summary of the prediction models, their stages of development, the cancer sites covered and the study designs 25

TABLE 10 Systematic review 2: risk-of-bias assessment of the included models 30

TABLE 11 Updated review: summary results from narrative synthesis, by cancer 36

TABLE 12 Externally validated predictive model accuracy in UK patients for populating the decision model 49

TABLE 13 Overview of studies included 51

TABLE 14 Economic decision-analytic models: study characteristics 53

TABLE 15 Implicit average median time to referral under current standard practice 76

TABLE 16 Mean time to referral of CRC patients by strategy in the model: base-case analysis 77

TABLE 17 Dukes' stage transitions 77

TABLE 18 Disease stage at diagnosis of CRC patients by strategy in the model: base-case analysis 78

TABLE 19 Model parameter values used in the base-case analysis 79

TABLE 20 Key model characteristics of the de novo model and previous models 82

TABLE 21 Clinical model outcomes of each strategy for CRC patients aged 70 years vs. a FIT of 20 µg Hb/g faeces 83

TABLE 22 Economic model outcomes of each strategy for CRC patients aged ≥ 70 years vs. a FIT of 20 µg Hb/g faeces 84

TABLE 23 Mean costs and QALYs per patient aged 70 years suspected of having CRC (0.18% prevalence) vs. a FIT of 20 µg Hb/g faeces 84

TABLE 24 Breakdown of discounted costs and QALYs in suspected cancer population 85

TABLE 25 Scenario analyses 88

TABLE 26 Numbers of years in practice and number of consultation sessions worked per week 99

TABLE 27 General practice software 99

TABLE 28 Responses to questions asking about availability of risk assessment tools, and likely use of paper-based tools (mouse mat or flip chart) 100

TABLE 29 Responses to question asking about eCDS tool availability 101

TABLE 30 General practices with access to the electronic clinical decision support tools 102

TABLE 31 Regression analysis output: dependent variable sex- and age-adjusted urgent 2WW referrals per 100,000 population 103

TABLE 32 Regression analysis output: dependent variable sex- and age-adjusted conversion rate (i.e. proportion of 2WW referrals that result in a diagnosis of cancer) 103

TABLE 33 Systematic review 1: included studies, study characteristics 147

TABLE 34 Systematic review 1: included studies, study design 148

TABLE 35 Systematic review 1: included studies, results 150

TABLE 36 Systematic review 1: included studies, critical appraisal 152

TABLE 37 Systematic review 2: included studies, study characteristics (1) 154

TABLE 38 Systematic review 2: included studies, study characteristics (2) 159

TABLE 39 Systematic review 2: included studies, model development and performance 178

TABLE 40 Systematic review 2: included studies, results 187

TABLE 41 Systematic review 2: risk-of-bias assessment, questions 1–3 201

TABLE 42 Systematic review 2: risk-of-bias assessment, questions 4 and 5 203

TABLE 43 Studies excluded from the review but highlighted for interest: studies with secondary care (i.e. referred or diagnosed) populations 205

TABLE 44 Studies excluded from the review but highlighted for interest: studies not on prediction models 206

TABLE 45 Systematic review 2 (updated searches): included studies, study characteristics (1) 207

TABLE 46 Systematic review 2 (updated searches): included studies, study characteristics (2) **207**

TABLE 47 Systematic review 2 (updated searches): included studies, model development and performance **208**

TABLE 48 Systematic review 2 (updated searches): included studies, results **208**

TABLE 49 Systematic review 2 (updated searches): risk-of-bias assessment, questions 1–3 **209**

TABLE 50 Systematic review 2 (updated searches): risk-of-bias assessment, questions 4 and 5 **209**

TABLE 51 Systematic review 2: updated searches **212**

TABLE 52 Description of studies included in the updated review **232**

TABLE 53 Updated review: results of included studies **242**

TABLE 54 Updated review: categories used for coding study aim **251**

TABLE 55 Updated review: study characteristics for the included CRC studies **252**

TABLE 56 Updated review: critical appraisal for the included CRC studies **261**

TABLE 57 Updated review: results of CRC studies **262**

TABLE 58 Summary results from previous review by Neal *et al.* **272**

TABLE 59 Conversion of Dukes' stages into TNM **294**

TABLE 60 Sensitivity analysis, parameter specification and ICER variation **296**

TABLE 61 Critical appraisal of model-based economic evaluation, Philips *et al.* **304**

TABLE 62 Probabilities and costs of adverse events of diagnostic tests in included studies **310**

TABLE 63 Assumptions in Westwood *et al.*'s model vs. assumptions in our model **317**

TABLE 64 Mean time to referral of CRC patients by strategy in the model **319**

TABLE 65 Disease state at diagnosis of CRC patients by strategy in the model **320**

TABLE 66 Parameters varied in probabilistic sensitivity analysis, the base case was for a 70-year-old individual and their distributions **320**

List of figures

FIGURE 1 The PRISMA flow diagram of the included studies for SR1 and SR2 **10**

FIGURE 2 Illustration of the different interval types **31**

FIGURE 3 Updated review: the PRISMA flow diagram **36**

FIGURE 4 Economic decision-analytic models: the PRISMA flow diagram **56**

FIGURE 5 Diagnosis pathway model **69**

FIGURE 6 Diagnostic decision model for patients with CRC **71**

FIGURE 7 Diagnostic decision model for patients without CRC **72**

FIGURE 8 Markov model of health-state transitions for CRC cases **74**

FIGURE 9 Relative survival of CRC patients diagnosed in 1996–2002 in England **78**

FIGURE 10 Actual and counterfactual survival curves for a patient who is diagnosed at Dukes' stage B **78**

FIGURE 11 Sensitivity of results to arbitrary increases and decreases of 20% in model parameter values **86**

FIGURE 12 Cost-effectiveness plane: diagnostic tool (QCancer) vs. no tool **86**

FIGURE 13 Cost-effectiveness acceptability curve: tool (QCancer) vs. no tool **87**

FIGURE 14 Threshold analysis **87**

FIGURE 15 Index of Multiple Deprivation profiles **100**

List of boxes

BOX 1 Updated review: bias assessment tool **258**

List of abbreviations

AJCC	American Joint Committee on Cancer
BB	Bristol–Birmingham
CAPER	Cancer Prediction in Exeter
CDSR	Cochrane Database of Systematic Reviews
CENTRAL	Cochrane Central Register of Controlled Trials
CI	confidence interval
CPRD	Clinical Practice Research Datalink
CRC	colorectal cancer
CTC	computerised tomography colonography
eCDS	electronic clinical decision support
FIT	faecal immunochemical test
FOBT	faecal occult blood test
gFOBT	guacal faecal occult blood test
GP	general practitioner
Hb	haemoglobin
HMIC	Health Management Information Consortium
HNPCC	hereditary non-polyposis colorectal carcinoma
HTA	Health Technology Assessment
ICER	incremental cost-effectiveness ratio
iFOBT	immunochemical faecal occult blood test
IMD	Index of Multiple Deprivation
IQR	interquartile range
IT	information technology
MeSH	medical subject heading
MISCAN	MIcrosimulation SCreening ANalysis
NAEDI	National Awareness and Early Diagnosis Initiative
NHS EED	NHS Economic Evaluation Database
NICE	National Institute for Health and Care Excellence
OTA	Office of Technology Assessment
PPI	patient and public involvement
PPV	positive predictive value
PROBAST	Prediction model Risk Of Bias ASsessment Tool
QALY	quality-adjusted life-year
RAT	risk assessment tool
RCT	randomised controlled trial
SEER	Surveillance, Epidemiology, and End Results
SimCRC	Simulation Model of Colorectal Cancer
SR	systematic review
TDI	total diagnostic interval
THIN	The Health Improvement Network
TNM	tumour node metastasis
WW	week wait

Plain English summary

In the UK, people with cancer tend to die sooner than people with cancer in other European countries. This may be because their cancers are caught at a later stage, perhaps after they have spread. Spotting cancer earlier in people, and testing them sooner, may extend people's lives. Researchers have developed 'diagnostic tools', which give the probability of having cancer, based on a patient's symptoms, blood test results and other information. The tools help family doctors decide who needs further testing for possible cancer, including cancers of the digestive, urinary and reproductive systems, and in the blood. We do not know how many family doctors have these tools, or how well the tools work.

We systematically reviewed published studies about how these tools were developed, how good and accurate they are, and what effects their use has on patients. We found that many tools have been developed, but there is little evidence that they improve the quality or length of life. We sent surveys to family doctors all over the UK asking if they had the tools at their practice and if they used them. Based on the replies we received, we estimate that the tools are in about one in three practices. They are likely to be used in about half of the practices where they are available. For practices in England only, we looked for, but did not find, any association between using the tools and the number of urgent appointments made for cancer testing.

We used a computer model to show what might happen if family doctors used the tools for patients who have symptoms of bowel cancer. In our model, if general practitioners used the tools, patients would need fewer appointments before they were referred to a specialist. This should reduce the time to diagnosis and treatment, compared with not using the tools. However, there is very little evidence as to whether or not this is indeed the case. Therefore, at the moment, we cannot say whether or not the use of such tools by general practitioners is better for patients and the NHS. More research is needed on what effect these tools have on patients, especially as to whether or not quality and length of life are improved.

Scientific summary

Background

Tools based on diagnostic prediction models are available to help general practitioners diagnose cancer. It is unclear whether or not they lead to increased or quicker diagnoses, and whether or not they ultimately affect patient quality of life and/or survival.

Objectives

The objectives were to evaluate the evidence on the validation, clinical effectiveness, cost-effectiveness (by two different systematic reviews), and availability and use of cancer diagnostic tools in primary care.

Systematic review 1

Methods

Two systematic reviews were conducted to examine the clinical effectiveness (systematic review 1) and development, validation and accuracy (systematic review 2) of diagnostic prediction models for use by general practitioners to aid cancer diagnosis. The following electronic databases were searched in May 2017 and updated in November 2018: MEDLINE, MEDLINE In-Process & Other Non-Indexed Citations, EMBASE, the Cochrane Library and Web of Science™ (Clarivate Analytics, Philadelphia, PA, USA). Titles, abstracts and full texts were screened independently by two reviewers.

Studies of any design were included in systematic review 1 if they assessed the clinical effectiveness of diagnostic tools in aiding decision-making among general practitioners for symptomatic patients presenting with features potentially indicative of cancer. An expanded definition of diagnostic tools was used, which included tools based on scoring systems/algorithms, as well as those based on prediction models.

Data extraction and assessment of risk of bias were completed by one reviewer and checked by a second reviewer. Owing to heterogeneity in tools, cancer sites, the outcomes measured and study design, a narrative review of the studies was conducted.

Results

Five studies met the inclusion criteria, and, between them, assessed three diagnostic tools: the risk assessment tools (as part of an education resource card in an Australian randomised controlled trial for lung, colorectal and prostate cancer, and mouse mats and desktop flip charts about colorectal and lung cancer in a UK-based pre–post study), a skin cancer algorithm (in a randomised controlled trial and a field trial, both based in Australia), and an online skin cancer recognition toolkit (in a UK-based case–control study).

Although the field trial and pre–post study reported a positive impact of the tools on outcomes, the results of the randomised controlled trials and the case–control study found no evidence that use of the tools was associated with better outcomes.

There is currently very little good-quality evidence to suggest that these tools can help improve general practitioner decision-making.

Systematic review 2

Methods

The search strategy was the same as that for systematic review 1. Studies of any design were included if they contained details on the development, validation or accuracy of diagnostic prediction models. Data extraction and assessment of risk of bias were completed by one reviewer and checked by a second reviewer. Owing to the heterogeneity of the tools, the cancer sites, the outcomes measured and the study design, a narrative review of the studies was conducted.

Results

A total of 43 studies met the inclusion criteria, including two systematic reviews. The searches identified evidence on 11 different prediction models in total, including risk assessment tools for 15 different cancer sites and QCancer® (ClinRisk Ltd, Leeds, UK) for six cancer sites, plus male and female versions for multiple cancers. Prediction models exist for 14 cancer sites, including models for multiple cancers. Colorectal cancer was associated with the greatest number of models ($n = 6$). The majority of QCancer models, one risk assessment tool and five other models have been externally validated.

There are clear gaps in the evidence for further validation of existing models that have the potential to be implemented in primary care to aid general practitioner decision-making.

Updated review

Methods

A review was conducted to update the findings of a previous systematic review that examined the association between different durations of time from first symptom to diagnosis or treatment, and clinical outcomes, across all major cancers. The updated review was conducted to inform the decision-analytic model and its structural assumptions. It therefore includes a more focused review of colorectal cancer.

Results

The updated review identified 35 new studies, the overall findings of which were summarised in a table outlining whether each study reported a 'positive association' (i.e. statistically significant more favourable patient outcomes), a 'negative association' (i.e. statistically significant less favourable outcomes) or 'no association' (i.e. the findings were not statically significant).

A more in-depth evaluation was conducted of colorectal cancer, which focused on studies identified during the updated review ($n = 10$) and better-quality studies identified in the previous review ($n = 4$). No meta-analyses were undertaken because of heterogeneity, which included variability in the intervals.

The majority of the colorectal cancer studies found 'no association' between various intervals and patient outcomes. A small number of studies ($n = 4$, but three used the same, or an overlapping, population) reported a positive association between shorter intervals and patient outcomes, but, paradoxically, a small number of studies ($n = 3$) also found a negative association.

These overall findings may reflect the U-shaped relationship between diagnostic interval and patient outcomes that was identified by some of the included studies, showing that both very short and long intervals were associated with poor outcomes. The review also identified important biases and other factors that may affect the findings of studies in this field.

Data for informing the economic decision model

Methods

The search strategy was designed to retrieve economic decision models for diagnosing or screening colorectal cancer. Colorectal cancer was the chosen focus for the economic analysis because its disease history in the UK setting has been researched in recent years. The methodological quality of the studies included was assessed in detail by two reviewers following the checklist for model studies by Philips *et al.* (Philips Z, Bojke L, Sculpher M, Claxton K, Golder S. Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. *PharmacoEconomics* 2006;**24**:355–71). Data extraction and assessment of risk of bias were completed by one reviewer and checked by a second reviewer. A narrative review of the studies was conducted.

Results

The searches identified 18 studies that met the inclusion criteria, which were then included in the review.

Our review found no evidence on the cost-effectiveness of diagnostic tools for managing patients in primary care with suspected colorectal cancer, but identified one study of faecal immunochemical tests in the low-risk population of interest that modelled the diagnostic phase. Our critique of the model identified shortcomings in the way time to referral and mortality were analysed in the diagnostic phase, which were to be addressed in the de novo model developed in the present study.

Economic decision model

Methods

A simple analytical model of diagnostic pathway was used to illustrate the uncertainty inherent in the current evidence base, and to ask questions about the probable impact of the diagnostic tools, given the current evidence base.

The model takes as its starting point symptomatic patients presenting to primary care who undergo an initial clinical assessment. This model is then combined with an adaptation of an existing disease model from a published colorectal cancer screening study and used to identify the parameters contributing most to the overall decision uncertainty about the cost-effectiveness of decision tools, and where additional research might be targeted in the future. In the absence of evidence on the impact of the tools on the time to diagnosis, a structural assumption was used to link the sensitivity of diagnostic strategies with the expected duration of the referral interval. The mechanism of effect of all the strategies considered in the model is, therefore, a reduction in the time to diagnosis, made possible by a reduction in the referral interval.

Results

The analysis using the limited available data on current practice in the UK suggests that the survival benefit of faster referrals for cancer patients is higher than the risks associated with exposing the overwhelming majority of patients without cancer to colonoscopy. Given the uncertainty in the evidence base, it is unclear if the overall benefits are worth the additional health-care costs associated with those referrals.

The sensitivity and threshold analysis revealed that the cost-effectiveness results were particularly sensitive to uncertainty around the diagnostic accuracy of current standard practice and the specificity of the tools. Other areas of uncertainty highlighted by the model include the clinical effectiveness of the tools, the prevalence of cancer in the low-risk population for which these tools are intended, the cost of colonoscopy and the definition of current practice.

General practice survey

Methods

A cross-sectional postal survey was carried out to determine (1) the proportions of UK general practices and UK general practitioners with access to cancer decision support tools and (2) the proportion of general practices that use cancer decision support tools. Data collection occurred in July and August 2017. Questionnaires were posted to 4600 general practitioners in 975 randomly selected UK practices. Using data from general practices in England only, ordinary least squares regression subanalyses explored the association between access to cancer decision support tools and practice-level cancer diagnostic indicators published by Public Health England. Ethics approval was granted by the University of Exeter.

Results

Responses were received from 473 general practitioners and three registrars in 227 practices, giving response rates of 23.3% (practice level) and 10.3% (practitioner level). Responding practices had a median of 6 (interquartile range 4–8) general practitioners, of whom a median of 2 (interquartile range 1–3) responded to the survey. EMIS Web (EMIS Health, Leeds, UK) was the most frequently used software (96/227, 42.3%), followed by TPP SystmOne (The Phoenix Partnership, Leeds, UK) (74/227, 32.6%) and then INPS Vision (In Practice Systems Ltd, London, UK) (32/227, 14.1%).

A total of 112 of the 476 general practitioners (23.5%, 95% confidence interval 19.7% to 27.6%) had access to a cancer decision support tool in either paper or electronic format, or both. At the practice level, at least one general practitioner in 83 of the 227 practices (36.6%, 95% confidence interval 30.3% to 43.1%) had access to a tool. Tools were available and likely to be used in 38 of the 227 practices (16.7%, 95% confidence interval 12.1% to 22.2%).

There was no difference in the mean 2-week-wait referral rate between practices that do and practices that do not have access to either type of tool, after adjusting for Index of Multiple Deprivation (mean difference 1.8 referrals per 100,000, 95% confidence interval −6.7 to 10.3). Access to either type of tool was not associated with a change in the proportion of 2-week-wait referrals that resulted in a diagnosis of cancer, after adjusting for the Index of Multiple Deprivation (mean difference −0.2, 95% confidence interval −1.0 to 0.6).

Discussion

Cancer decision support tools are available to general practitioners in approximately one-third of UK general practices, but are likely to be used in only one-sixth of practices.

Improvements in training and increasing familiarisation with the tool may increase the levels of uptake of these tools by UK general practices and general practitioners.

More research is needed to determine the comparative accuracy of the tools in studies that directly compare them with current standard practice and in the same low-risk suspected symptomatic patient population in primary care. To inform decisions about the use of the tools to aid diagnosis in primary care, such studies should aim to measure the impact of the tools on diagnostic intervals and, ideally, on clinical outcomes.

Conclusions

Our survey indicates that cancer decision support tools are currently not widely used in the UK. This may reflect our findings in systematic reviews 1 and 2 that there is limited evidence that these tools have a positive impact on patient outcomes.

As levels of uptake are currently low, it is possible to carry out a randomised controlled trial to assess whether or not these tools are genuinely helpful in improving the selection of patients for investigation for suspected cancer.

Study registration

This study is registered as PROSPERO CRD42017068373 and CRD42017068375.

Funding

This project was funded by the National Institute for Health Research (NIHR) Health Technology programme and will be published in full in *Health Technology Assessment*; Vol. 24, No. 66. See the NIHR Journals Library website for further project information.

Chapter 1 Introduction

Scientific background and rationale

The rate of cancer survival in the UK is lower than the European average for most cancers: for example, the 5-year survival rate for stomach cancer is 17.2% in the UK, compared with the European average of 25.1%; for colon cancer, it is 51.8% in the UK, compared with the European average of 57%.[1] Efforts to reduce the time to making a cancer diagnosis have the potential to improve prognosis,[2] because earlier diagnosis is associated with earlier cancer stage at diagnosis,[3] and earlier treatment is associated with improved survival.[4] There is also the potential to reduce presentation via emergency admissions, and to prevent the poorer survival associated with that route of diagnosis.[5] National cancer screening programmes in the NHS (for breast, bowel and cervical cancer) and the National Awareness and Early Diagnosis Initiative (NAEDI) (to increase public awareness of the signs and symptoms of cancer[4]) are intended to improve early diagnosis. As many individuals go through primary care as a route for diagnosis,[5] efforts there could improve cancer survival.

Cancer diagnosis in primary care is not straightforward. Symptoms of cancer are commonly seen, but mostly have non-cancer origins.[6] Of those individuals referred from primary care via the 2-week-wait (2WW) referrals for suspected head and neck cancer, approximately 9% were ultimately diagnosed with cancer.[7] The type and presence of symptoms can vary greatly,[8] 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.[9] Thus, tools to help improve cancer diagnosis in primary care have great potential to affect 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 in an individual patient.[10,11] These prediction models can then be used to develop diagnostic tools (such as a website risk calculator or a mouse mat detailing estimates of risk depending on features) to assist doctors in estimating probabilities, and can potentially influence doctors' decision-making.[11] 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.[12] 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. In the literature on prediction models in general, very few diagnostic prediction models that are developed go on to be evaluated for their clinical impact.[12]

Tools currently available to GPs to help cancer diagnosis, beyond the National Institute for Health and Care Excellence (NICE) guidelines for suspected cancer referral,[6] are based on the following diagnostic prediction models:

1. the risk assessment tool (RAT) developed by Hamilton *et al.*,[13] which provides estimates of cancer risk for 17 cancers, based on symptoms alone
2. the QCancer® (ClinRisk Ltd, Leeds, UK) tool, which estimates the risk of 10 cancers, based on symptoms and patient characteristics, such as age, smoking status and body mass index.

There are clear differences in the derivation of the RAT and QCancer. The RAT used a case–control design to predict likely cancer diagnosis, whereas QCancer used a cohort design. Many of the QCancer prediction models have subsequently been externally validated and reported to have good diagnostic performance.[14,15] There has, however, been no comparison of the clinical effectiveness of these diagnostic tools in clinical practice, or research on whether or not GPs currently have access to and are using these tools.

In 2013, Hamilton *et al.*[13] reported an increase in cancer referrals and investigations associated with the introduction of RATs as mouse mats and desktop flip charts for lung cancer and colorectal cancer (CRC), and an increase in the awareness of GPs of cancer symptoms, especially those symptoms that are less known in those cancers.[16] A 2015 evaluation of an electronic version of RATs for lung cancer and CRC highlighted the potential issue of prompt overload from the system, cautioned on potential variation in data used by the tool, and the extent to which the aid might increase pressure on secondary care owing to increased referral (a finding that could be generalised to all such diagnostic tools).[17]

An Australian study using simulated GP consultations explored the implementation of an aid based on QCancer.[18] The study found that GPs agreed that the diagnostic aid was potentially useful in practice, but noted that different GPs interpreted the same set of symptoms differently, leading to inconsistent estimates of risk from the QCancer aid. In collaboration with the NAEDI, Macmillan Cancer Support developed, with BMJ Informatica, and evaluated the introduction of an electronic clinical decision support (eCDS) containing the RAT and QCancer for colorectal, lung, oesophagogastric, pancreatic and ovarian cancers. It was found that the impact of eCDSs varied across practice, from no impact on referrals to increased referrals and investigations in other practices,[19] with use of eCDSs leading to further investigation or referral of the patient that would not have occurred otherwise in 19% of cases.

However, there is very little evidence as to whether or not these tools have led to increased or quicker cancer diagnoses, and, ultimately, to impacts on patient quality of life or survival. A study protocol for a randomised controlled trial (RCT) to evaluate eCDSs for assessing symptoms indicative of stomach cancer was published in 2016.[20] Other diagnostic prediction models have been developed in the UK, such as that reported by Iyen-Omofoman *et al.*[21] for lung cancer and the Bristol–Birmingham (BB) equation for CRC,[22] plus those developed outside the UK, such as Benign, Lonely, Irregular, Nervous, Change, Known (BLINCK) clues in Australia for skin cancer[23] and that developed in the USA for ovarian cancer,[24] which may have the potential to be useful in the NHS context. However, little is known about whether or not, and how, these diagnostic prediction models and tools affect patient outcomes, and would affect NHS resources.

Although we are unclear about the evidence on the clinical effectiveness of these diagnostic tools to affect patient quality of life and survival, a systematic review conducted by Neal *et al.*[25] found a large number of studies looking at the impact on patient outcomes of reducing diagnostic and/or treatment intervals for cancer. Only a small number of studies were found to be of high quality, and there was substantial variation in the type of intervals evaluated and the findings within and between cancer types. Compared with other cancer types, studies of colorectal, breast, head and neck, and testicular cancers and melanoma suggested that shorter time intervals were associated with improved patient outcomes. However, for each of these cancer types, there were also studies reporting no association between time interval and patient outcome.

The possible trade-offs between the costs and the harms, and the inherent uncertainty, of using these diagnostic tools in primary care are also unclear, as well as the extent to which reducing times to diagnosis and/or treatment could affect patient outcomes.

Aims and objectives

The aim of this project was to evaluate the evidence on the development, validation, clinical effectiveness and cost-effectiveness of cancer diagnostic tools in primary care, and to understand the extent to which existing tools are currently used in the primary care setting in the NHS.

The objectives were to:

1. identify evidence evaluating the clinical effectiveness of symptom-based diagnostic tools that that could be used to inform cancer diagnosis decision-making in primary care (see *Chapters 2* and *3*)
2. identify and summarise studies reporting the development, validation or accuracy of any diagnostic prediction model that could be used as a tool to aid cancer diagnosis in primary care (see *Chapters 2* and *4*)
3. update a previous systematic review[25] assessing the association of the durations of different intervals in the diagnostic process to clinical outcomes, and conduct a more in-depth evaluation of CRC studies, focusing on methods, to identify studies that are likely to provide the best estimate of the impact of diagnostic intervals on patient outcomes for informing the decision-analytic mode (see *Chapter 5*)
4. use a decision-analytic model to explore uncertainties in the cost-effectiveness of using symptom-based diagnostic tools, including the impacts on health service resource use, costs and patient outcomes, using CRC as an example (see *Chapters 6* and *7*)
5. understand the extent to which GPs currently have access to cancer diagnostic tools and are using them in primary care to inform their decision-making (see *Chapter 8*).

Chapter 2 Systematic reviews 1 and 2: literature search strategy

Introduction

Two systematic reviews were conducted: systematic review (SR) 1 and SR2.

The research question (SR1) was 'what evidence is there for the clinical effectiveness and cost-effectiveness of symptom-based diagnostic tools that could be used to inform cancer diagnosis decision-making in primary care?'. It was anticipated that limited evidence would be identified; therefore, a second review was planned (SR2) to identify studies reporting the development and validation of any diagnostic prediction model that could be used as a tool to help cancer decision-making in primary care, that is to provide a list of cancer models that might have the potential to be developed into diagnostic tools. This chapter clarifies the definitions of diagnostic tools and prediction models used and describes the search methods employed for SR1 and SR2. The results are presented separately in *Chapters 3* (SR1) and *4* (SR2).

Diagnostic prediction models can be categorised according to the different stages of development and evaluation of the model (*Table 1*). However, a number of predictive research studies have differed in what they consider to fall under the 'predictive model research' header. This variation primarily related to whether or not they incorporated predictor finding studies as an initial stage and, at the other end of the spectrum, whether or not they incorporated impact or implementation studies. For the purpose of our reviews, we have differentiated between prediction models and prediction tools. 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.[26] Symptoms could be self-reported by the patient, or prompted by a physician's questioning. Signs and test results are identified in primary care via routine testing (e.g. full blood count, urine dipstick testing, clinical signs), and patient characteristics are also determined in primary care (e.g. sociodemographic variables, personal and family history). The prediction tools implement the models to provide a numerical risk of having cancer. As examples, prediction tools may be mouse mats or desktop flip charts, or may be integrated into information technology (IT) systems.

As a 2009 study by Moons *et al.*[28] points out, studies assessing the impact of prediction tools need designs and outcome measures that are different from those used in studies that develop and evaluate prediction models. Studies of predictive research[26,27,29,30] also differ slightly in the way they categorise studies, based on whether they considered the evaluation of internal validity to be part of the model development stage or a separate stage.

Table 1 summarises the spectrum of potential study types and clarifies the inclusion and exclusion criteria for SR1 and SR2.

Methods

The systematic reviews were conducted in accordance with good practice guidelines.[31] As the inclusion and exclusion criteria were very similar for SR1 and SR2, the same search strategy was used for both reviews; however, two separate protocols were developed for reviewing the evidence. Further details are presented in this chapter.

TABLE 1 Stages of development of a diagnostic prediction model

Predictive research stage	Classification of stage used for coding studies	Description	Exclude or include
Identifying single predictors	Predictor identification	Studies that aim to explore which predictors out of a number of candidate predictors independently contribute to the prediction of (i.e. are associated with) a diagnostic (or prognostic) outcome[26]	Exclude
Model development only	Development only (apparent performance)	Studies in which performance is directly evaluated using exactly the same data used to derive the model	Include in SR2
Model development with internal validation	Internal validation I	Studies that use only the original study sample for both development and validation using resampling techniques (e.g. cross-validating, bootstrapping, jackknifing)	
	Internal validation II	Studies that use only the original study sample for both development and validation using split sampling, whereby part of the sample is used for model derivation and the other part is used for validation	
External validation	External validation	Studies that aim to assess and compare the predictive performance of an existing prediction model using new participant data that were not used in the development process	
Model updating	Model update	External validation studies in which the model is adjusted or updated in the case of poor performance, based on the validation	
Impact assessment	Impact assessment	Studies that aim to quantify the effect or impact of using a diagnostic tool (relative to not using the tool) on patient or physician behaviour and management, patient health outcomes, or cost-effectiveness of care.[26] There are two types of analyses: 1. Narrow impact analysis – prospective demonstration in one setting that use of the prediction rule improves physicians' decisions (quality or cost-effectiveness of patient care), which could be used to inform decisions in similar settings[27] 2. Broad impact analysis – prospective demonstration in varied settings that use of the prediction rule improves physicians' decisions for a wide spectrum of patients[27]	Include in SR1

Search strategy

Bibliographic searches of relevant databases [MEDLINE (1946 to May week 1 2017), MEDLINE In-Process & Other Non-Indexed Citations, EMBASE (1974 to 10 May 2017), the Cochrane Library and Web of Science™ (Clarivate Analytics, Philadelphia, PA, USA)] were conducted in May 2017, with updated searches conducted in November 2018. SR1 and SR2 were conducted in parallel using the same search strategy, but each review had different inclusion and exclusion criteria.

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 (*Table 2*). No date, language, study design or other limits were used. Search filters for clinical prediction models were

TABLE 2 Search strategy for MEDLINE (searched May 2017)

1.	exp Neoplasms/
2.	(cancer$ or neopla$).tw.
3.	(tumour$ or tumor$).tw.
4.	or/1-3
5.	Primary Health Care/
6.	exp General Practice/
7.	General Practitioners/
8.	(primary care or general practi$ or family practi$).tw.
9.	(primary adj3 (healthcare or health care)).tw.
10.	Or 5/9
11.	Decision Support Systems, Clinical/
12.	Decision Support Techniques/
13.	(tool or tools or aid$ or model or models or checklist$ or check list$ or rule or rules or algorithm$ or equation$).tw.
14.	or/11-13
15.	"Early Detection of Cancer"/
16.	(predict$ or assess$ or scor$ or risk$ or validat$ or decision$ or identif$ or diagno$ or prognos$).tw.
17.	(2ww or 2 week wait or two week wait or 2 week rule or two week rule).tw.
18.	or/15-17
19.	4 and 10 and 14 and 18

investigated but none was thought to be fully tested or reliable. A balance was sought between the sensitivity of the search results and the number of papers to be screened.

The search results were exported to EndNote X7 [Clarivate Analytics (formerly Thomson Reuters), Philadelphia, PA, USA] and de-duplicated using automatic and manual checking.

Items included after full-text screening were forward and backward citation-chased using Scopus® (Elsevier, Amsterdam, the Netherlands) to identify additional relevant studies. Once relevant models and tools were identified from the initial searches, additional searches were conducted to identify names of tools [e.g. QCancer, RAT, Cancer Prediction in Exeter (CAPER), BB equation, GP Skin Cancer Toolkit], to ensure that search results were sufficiently comprehensive.

The full search strategies and results are in *Appendix 1*.

Inclusion and exclusion criteria

For SR1, *diagnostic tools* were considered initially as *diagnostic prediction models* that are used in clinical practice to assist doctors in estimating probabilities to aid decision-making. Pilot searches identified one study that assessed the impact of implementing diagnostic prediction models. Therefore, the definition of 'diagnostic tool' for SR1 was expanded to include any quantitative tool used to support a GP in deciding which patient warrants further investigation for cancer. Such investigation could be via referral to secondary care or involve further testing in primary care. In other words, the diagnostic tools may be based not only on diagnostic prediction models, but also on scoring systems/algorithms, etc. Studies were excluded that simply looked at 'red-flag symptoms' or symptom lists and (weighted)

scores that did not provide a numerical risk of current cancer. Owing to the limited anticipated number of relevant studies, we sought any study reporting on impact, regardless of study design.

For SR2, 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.[26] Symptoms could be self-reported by the patient or prompted by a physician's questioning. Signs and test results are identified in primary care via routine testing (e.g. full blood count, urine dipstick testing, clinical signs), and patient characteristics are also determined in primary care (e.g. sociodemographic 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 included only if an attempt was made to validate the models with primary care data.

Inclusion and exclusion criteria for SR1 and SR2 are presented in *Table 3*.

Selection of studies

Although slightly different inclusion and exclusion criteria were used for SR1 and SR2, the screening of articles was conducted simultaneously by two reviewers (RL and BG). An algorithm was used whereby if studies met the mutual inclusion criteria for SR1 and SR2 (population, setting, publication type), they were then assessed further as to whether they were appropriate for SR1 or SR2 (or excluded).

Titles and abstracts were screened for relevance independently; any disagreements were resolved by consensus. Pilot screening was undertaken for the first 100 hits to ensure that 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 (RL and BG). Disagreements were discussed between the two reviewers; if not resolved, a third reviewer (Christopher Hyde) made the final decision.

For SR2, multiple studies reported the development and validation aspects of particular prediction models (e.g. the development and internal validation of the prediction model by Hippisley-Cox *et al.*[32] in one paper, and the external validation in a separate paper.[14] All studies related to each specific prediction model were collated, regardless of whether they refer to the development or validation of that tool.

Results

Studies identified

Search phrases were finalised and searches were run in May 2017 (see *Appendix 1*). 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 January 2018, resulting in 631 additional new records (after de-duplication), and again in November 2018, when 702 hits were identifeid. Discussions with collaborators led to the identification of relevant grey literature, but no such studies were deemed eligible for inclusion.

As hits were screened simultaneously for inclusion in SR1 or SR2, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)[33] flow diagram shows the results for both reviews: 260 full-text articles were screened: five studies met the inclusion criteria for SR1 and 41 studies met the inclusion criteria for SR2 (*Figure 1*).

Details on the methods for data extraction, on assessing the risk of bias of the included studies and on the findings of the SRs are presented in *Chapters 3* (SR1) and *4* (SR2).

TABLE 3 Inclusion and exclusion criteria

Criterion	SR1	SR2
Population	Included: symptomatic patients (with symptoms being indicative of cancer) presenting at primary care or patients referred with symptoms indicative of cancer	
	Excluded: asymptomatic patients (screening population)	
Technology	Included: feature[a]-based diagnostic tools implemented/used in primary care to provide additional information on the risk of cancer. The tool may be used for the purpose of diagnosing cancer in primary care (leading to referral for treatment) or to inform decisions about referring for further tests (with possible diagnosis occurring in secondary care)	Included: diagnostic prediction models, based on two or more features,[a] that estimate the risk of prevalent but undiagnosed cancer
	Excluded: prognostic or screening prediction models; statistical tools that estimate the probability of developing cancer over a defined period of time	
Setting	Included: primary care	
	Excluded: secondary care; online tools developed for use by the general population	Exclusion: models developed into tools for online use by the general population
Study design	Included: comparative studies of diagnostic tools that assessed impact in clinical practice (RCTs, controlled before and after, and interrupted time series); studies analysing national trends in cancer diagnosis before and after diagnostic tools became available	Included: any design for the development, validation or accuracy of diagnostic prediction models (as defined in the 'Technology' row of this table)
	Excluded: uncontrolled studies reporting qualitative data	
Comparison	Usual care or the use of another diagnostic tool	N/A

Outcomes	Primary outcomes:
	• 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)
	• Survival
	• Economic outcome measures (resource use, cost per diagnosis, cost per QALY)
		• Estimates of the risk of being diagnosed with cancer (e.g. ORs, HRs)
		AND/OR
		• Any details on the development, validation or accuracy of the tool:
		  ○ Model development – method, assumptions, predictors, shrinkage, coefficient weighting
		  ○ Model evaluation (validation)
		  ○ Assessing (quantifying) model performance – 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. R^2), classification (e.g. sensitivity, specificity, predictive values)

	Secondary outcome: • Referral patterns	Excluded: models that report the risk of survival (or stage at diagnosis, etc.)
Publication type	Included: published in full and English-language publication	
	Excluded: commentaries, letters	

HR, hazard ratio; N/A, not applicable; OR, odds ratio; QALY, quality-adjusted life-year; ROC, receiver operating characteristic.
a Features include symptoms and other information, such as elicited signs, patient characteristics and test results.

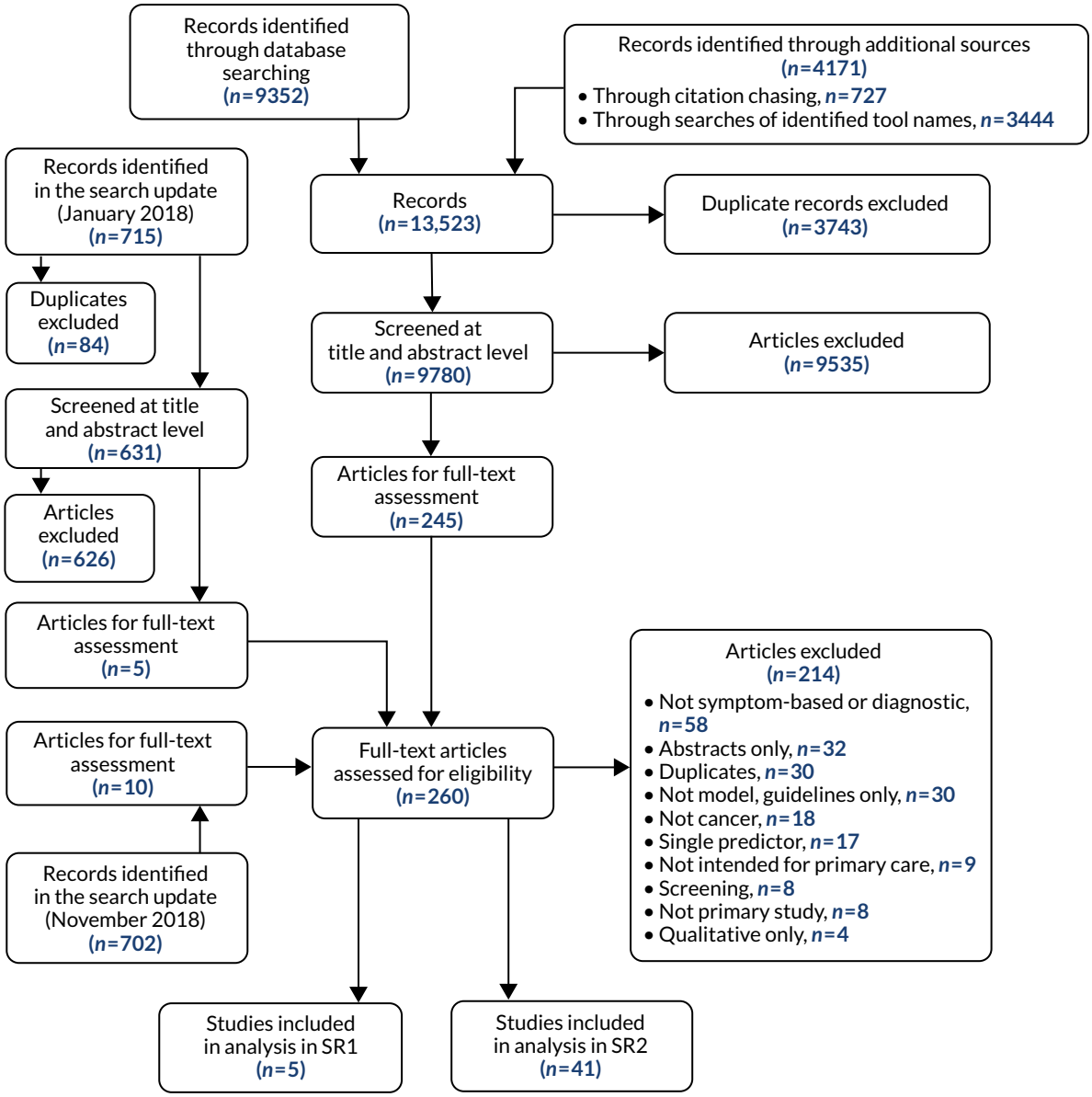


FIGURE 1 The PRISMA flow diagram of the included studies for SR1 and SR2.

Chapter 3 Systematic review 1

Objective

The objective was to identify evidence evaluating the clinical effectiveness of symptom-based diagnostic tools that that could be used to inform cancer diagnosis decision-making in primary care.

Methods

Identification of studies

Information related to the search strategy, eligibility criteria and selection of studies is provided in *Chapter 2*.

Data extraction

To extract relevant data from each included study, standardised data extraction forms were used, which evolved following piloting and discussion among reviewers. One reviewer (BG) extracted the data, which were checked by a second reviewer (RL). Extracted data included cancer type(s); study design; country; sample size; patient recruitment (with inclusion and exclusion criteria); characteristics of the tool (including whether based on symptoms alone or other features in addition to symptoms); definition of outcomes (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); main results, including confidence intervals (CIs); and subgroup analyses, when available.

Critical appraisal

A risk-of-bias form based on the Cochrane Effective Practice and Organisation of Care group recommendations[34] was used to assess potential features of different study designs that may lead to biased estimates of clinical effectiveness. This was conducted by one reviewer (BG) and checked by a second reviewer (RL).

Data synthesis

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

Results

Studies identified

The impact of three 'overarching' diagnostic tools was assessed by five studies:[13,35–38] an algorithm for differentiating malignant and benign skin lesions, a skin cancer toolkit and the RATs, which have been developed for various cancer sites. The data extracted from the five included studies are presented in *Appendix 2*.

Study design was heterogeneous, consisting of RCTs,[35,36] one field trial,[37] one cohort study[13] and one case–control study.[38]

Three of these studies[36–38] assessed two different decision support tools for skin cancer (*Table 4*). Del Mar and Green[37] and English *et al.*[36] describe two evaluations of an algorithm to improve the diagnosis of malignant melanocytic lesions and, consequently, reduce the proportion of total lesions excised that prove to be benign. Both studies were conducted in Australia. The Del Mar and Green[37] study was designed as a field trial in two cities: the impact of using the algorithm in one city was compared with not using it in the second city, which acted as a control.

Gulati *et al.*[38] evaluated the impact of a UK-wide, online, skin cancer recognition toolkit on GP confidence and knowledge in diagnosing skin cancers and referral behaviour. Additional skin cancer referral data were obtained to assess the appropriateness of referrals, and a survey was also conducted to investigate GP confidence in diagnosing skin cancer.

The other two included studies[13,35] evaluated the impact of previously developed diagnostic prediction models in practice. Both evaluated the use of RATs. Hamilton *et al.*[13] investigated the number of times two RATs[39] (one for lung cancer and one for CRC) 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 *et al.*[35] 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 with 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[39] (further details are provided in *Chapter 4*). Emery *et al.*[35] used the total diagnostic interval (TDI), that is the time from first symptom to cancer diagnosis, as an outcome measure.

TABLE 4 Systematic review 1: description of the tools assessed

Cancer type(s)	Prediction tool	Study	Country of tool development	Tool description
Skin cancer	Melanoma 'algorithm' (plus camera)	Del Mar 1995;[37] English 2003[36]	International (meta-analysis), adapted to Australian guidelines	An algorithm for managing clinically suspicious naevi, aided by the use of a camera
Skin cancer	GP skin cancer toolkit web resource	Gulati 2015[38]	International (meta-analysis), adapted to Australian guidelines	The toolkit consisted of a referral decision aid (referral guidelines based on red flags), lesion recognition resource (a series of images), clinical cases and a quiz
Multiple (lung, colorectal)	RAT presented on a mouse mat and desktop flip chart	Hamilton 2013[13]	UK	RAT gives risk estimates for patients aged > 40 years presenting to primary care with symptoms of possible cancer, for single symptoms, pairs of symptoms and repeat attendances with the same symptom. The values are colour coded to aid interpretation
Multiple (breast, prostate, colorectal or lung)	Education resource card containing the RAT	Emery 2017[35]	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's guidelines for investigating new breast symptoms

Critical appraisal

Studies were heterogeneous in how they addressed risk of bias (*Table 5*). Three of the studies were not randomised. Allocation blinding was another area of vulnerability for the majority of the studies, although this could not be assessed for one of them.[35] All studies managed to ensure reasonably similar baseline characteristics and measurements for the study groups. Among the studies included, Emery *et al.*[35] raised the fewest concerns for risk of bias.

Study outcomes

The outcome that Del Mar and Green[37] used was the percentage of lesions excised that were benign. The study[37] observed that use of the algorithm seemed to reduce the ratio of excised benign lesions to melanomas (from 93.8% to 88.8%; $p < 0.001$), without reducing the number of melanomas diagnosed. English *et al.*[36] used a 'slightly modified' algorithm that was evaluated by randomising general practices in Perth, Australia, to use the algorithm, while others were used as controls. The outcome used was the ratio of excised benign lesions to excised melanomas. The study[37] found no reduction in the ratio of benign to malignant lesions excised.

Gulati *et al.*[38] showed no significant changes in the number of urgent GP referrals for suspected skin cancer, diagnoses of melanoma or diagnoses of non-melanoma skin cancer between the toolkit users and the non-users in the study periods, despite increased GP confidence in making skin cancer referrals. The proportion of appropriate referrals increased with the use of the toolkit; however, the differences between toolkit users and non-users did not reach statistical significance.

Hamilton *et al.*[13] reported on changes in investigations carried out and rapid referrals before and after the introduction of the tools. They found an increase of 31% in rapid referrals for lung cancer and a 4% increase in GP-mandated chest X-ray investigations, as well as a 26% increase in referrals for CRC 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.

Emery *et al.*[35] 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 subintervals of the TDI.

None of the included studies reported outcomes such as diagnoses made, quality of life, survival or NHS resource use.

Study results are summarised in *Table 6*.

Discussion

This review attempted to summarise existing evaluations of feature-based cancer diagnostic tools used in primary care. Our strategy was able to identify a limited number of heterogeneous studies that did not provide strong evidence of the impact of feature-based diagnostic tools on patient-related outcomes or referral patterns. The small number of studies ($n = 5$) and the heterogeneity in reported outcomes did not allow for a meta-analysis of the results. The included studies provided limited evidence of the clinical effectiveness of using diagnostic tools.

A few other reviews have looked at feature-based cancer diagnostic tools in primary care. Williams *et al.*[40] conducted a systematic review of studies that described, validated or assessed the impact of CRC diagnostic tools. However, they did not identify any studies that tested whether or not patients who were diagnosed with the aid of the tool fared better than those who were diagnosed without it. Schmidt-Hansen *et al.*[41] conducted a similar review of lung cancer tools and found limited evidence to

TABLE 5 Systematic review 1: risk-of-bias assessment

Study	Bias								
	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
RCTs									
English 2003[36]	✓	✓	✓	✓	?	✗	✓	✓	?
Emery 2017[35]	✓	✗	✓	✓	✓	✓	✓	✓	✓
Field trials									
Del Mar 1995[37]	✗	✓	✓	✓	✗	✗	✓	✓	?
Case–control study									
Gulati 2015[38]	N/A	N/A	✓	✓	?	N/A	N/A	?	✗
Pre–post study									
Hamilton 2013[13]	N/A	N/A	N/A	N/A	?	N/A	N/A	?	✗

✓, low risk; ✗, high risk; ?, unclear risk; N/A, not applicable.

TABLE 6 Systematic review 1: results reported by the studies

Cancer type(s)	Study	Prediction tool	Country	Study design	Intended purpose	Main results
Melanoma	Del Mar 1995[37]	Melanoma 'algorithm' (plus camera)	Australia	Field trial	To evaluate whether or not an algorithm can reduce the number of benign lesions being excised without reducing the excision of invasive lesions, by comparing numbers of excised lesions with and the number without algorithm use	Total number of excised lesions • At baseline: control city, 752; intervention city, 606 • After intervention: control city, 2468; intervention city, 1997 Percentage of excised lesions that are neither invasive or potentially malignant • At baseline: control city, 94% (95% CI 92.3% to 95.7%); intervention city, 93.6% (95% CI 91.6% to 95.5%) ($p = 0.731$) • After intervention: control city, 93.8% (95% CI 92.8% to 94.8%); intervention city, 88.8% (95% CI 87.4% to 90.2%) The median number of excisions per doctor (2.5% and 97.5% percentiles) • At baseline: control city, 7 (1, 38); intervention city, 8 (1, 46) • After intervention: control city, 9 (1, 39); intervention city, 4 (1, 30) There were significant differences in the percentages of benign lesions reported in the intervention and control cities (88.8% and 93.8%; $p < 0.001$) after the intervention

continued

TABLE 6 Systematic review 1: results reported by the studies (*continued*)

Cancer type(s)	Study	Prediction tool	Country	Study design	Intended purpose	Main results
Melanoma	English 2003[36]	Melanoma 'algorithm' (plus camera)	Australia	RCT	To determine whether or not an aid to the diagnosis of pigmented skin lesions reduces the ratio of benign lesions to melanomas excised in general practice	Number of excised skin lesions (including seborrheic keratoses) At baseline: • Control – benign, 1965; melanoma, 61; ratio 32 • Intervention – benign, 2615; melanoma, 100; ratio 26 Trial period: • Control – benign, 2037; melanoma, 79; ratio 26 • Intervention – benign, 2369; melanoma, 81; ratio 29 ($p = 0.88$) Provision of the algorithm and camera did not decrease the ratio of benign pigmented skin lesions to melanomas excised by GPs (OR 1.03, 95% CI 0.71 to 1.50; $p = 0.88$)

Cancer type(s)	Study	Prediction tool	Country	Study design	Intended purpose	Main results
Skin cancer	Gulati 2015[38]	GP Skin Cancer Toolkit	UK	Case–control	To assess the impact of the toolkit by comparing before-and-after national skin cancer referral data, data from cross-sectional questionnaires and data on urgent skin cancer referrals to two NHS trusts	21,000 GPs were invited to use the tool; 8163 GPs accessed the tool during the 2012 period. There were no significant changes in the number of urgent GP referrals for suspected skin cancer (Spearman's rank 0.20; $p < 0.001$), diagnoses of melanoma (Spearman's rank 0.064; $p < 0.001$) or diagnoses of non-melanoma skin cancer (Spearman's rank 0.068; $p < 0.001$) between the toolkit user and the non-user groups. The proportion of appropriate referrals increased from 21.37% in 2011 to 32.3% in 2012, giving an incidence rate ratio of 3.13 (95% CI 2.21 to 4.42, z-statistic 6.46; $p < 0.0001$) The differences in numbers of appropriate referrals between toolkit users and non-toolkit users did not reach statistical significance by Spearman's rank test and ANOVA
Multiple (lung, colorectal)	Hamilton 2013[13]	RAT for lung cancer and CRC in two formats: mouse mat and desktop flip chart	UK	Pre–post study	To compare referrals and investigations for colorectal and lung cancer before and after the implementation of RATs	Lung cancer: 31% increase in 2-week referrals (332 before, 436 after); 4% increase in related investigations (chest X-ray) (7431 before, 7723 after) CRC: 26% increase in 2-week referrals (1173 before, 1477 after); 15% increase in colonoscopies (1762 before, 2032 after) No conclusion possible on the clinical effectiveness of the intervention

continued

TABLE 6 Systematic review 1: results reported by the studies (*continued*)

Cancer type(s)	Study	Prediction tool	Country	Study design	Intended purpose	Main results
Multiple (breast, prostate, colorectal or lung)	Emery 2017[35]	Education resource card including RAT for colorectal, lung and prostate cancers	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	No significant differences in the median or ln mean TDI at either intervention level: • GP intervention vs. control: median TDI 97 vs. 96.5 days; log-transformed mean difference 0.004 (95% CI −0.18 to 0.19; $p = 0.99$) • Community intervention vs. control: median TDI 107.5 vs. 92 days; log-transformed mean difference 0.08 (95% CI −0.06 to 0.23; $p = 0.27$) No significant differences in the TDI when analysed by factorial design, tumour group or subintervals of the TDI

ANOVA, analysis of variance; OR, odds ratio.

support the recommendation of any of the identified risk prediction tools, owing to lack of external validation or cost impact assessment. Similarly, Usher-Smith *et al.*[42] concluded that, even though some of the prediction models had the potential for clinical application, there remains considerable uncertainty about their clinical utility.

Other reviews have looked at cancer RATs in primary care.[43,44] However, they differ from this review in that they reported on tools that estimate the risk of developing cancer in the future, rather than the risk of having an undiagnosed cancer based on current signs and symptoms.

Although the intention of our review was to explore tools based only on prediction models, it was not clear whether or not, in practice, the impact of such tools could be isolated from other decision-making tools available to practitioners, such as diagnostic algorithms[45] or guidelines.[6] With limited evidence available on the impact of implemented diagnostic models, we decided to report on identified studies on the algorithm-based tools as well; however, the evidence was still sparse.

Among the limitations of the included studies were lack of randomisation, lack of patient-related outcomes and use of models developed on different populations. The outcome measures used by some of the 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 versus non-cancer diagnosis.

Furthermore, concerns on the quality of the studies make 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[35] also 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 clinical effectiveness of the diagnostic tools. These findings could be further obfuscated by publication bias, whose magnitude on this topic remains unknown.

Conclusion

Current evaluations provide limited evidence of the impact on patient outcomes of using feature-based cancer diagnostic tools in primary care. Better research is needed to provide these data, possibly through better study design and choice of outcomes. However, identifying the ideal approach 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.

Chapter 4 Systematic review 2

Objective

Systematic review 2 was conducted as a complementary study to SR1, described in *Chapter 3*. The objective was to identify and summarise studies reporting the development, validation or accuracy of any diagnostic prediction model that could be used as a tool to aid cancer diagnosis in primary care.

Methods

Identification of studies

Information related to the search strategy, eligibility criteria and selection of studies is provided in *Chapter 2*.

Data extraction

An adaptation of the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS)[30] was used to extract the following data from each included study: country, cancer type(s), study design, data source, sample size, number of participants with specific cancer, recruitment (including inclusion and exclusion criteria), participant characteristics, 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) and features included in final model.

Critical appraisal

Risk of bias was assessed with the use of a form based on the work of the Prediction model Risk Of Bias ASsessment Tool (PROBAST)[46] group. Because a final version of this checklist was not publicly available at the time of the appraisal, we followed recent recommendations on reporting reviews of prediction models.[47–49] Like the PROBAST checklist, the derived checklist assesses the risk of bias and applicability of prediction-modelling studies on five domains: participant selection, predictors, outcome, sample size and missing data, and analysis (*Table 7*).

Data synthesis

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

Results

Studies identified

There were 41 included records from the searches, including two systematic reviews. A further two studies were identified from one of the reviews. The primary are summarised in *Appendix 3*.

Systematic reviews

Two included records[41,50] are systematic reviews that had some overlap with the included studies described in this section (*Table 8*).

Schmidt-Hansen *et al.*[41] conducted a systematic review of the literature to identify risk prediction tools to be used in primary care to aid diagnosis of lung cancer. Five separate tools were identified: RAT,[51] QCancer,[52–54] the equation from Iyen-Omofoman *et al.*,[21] and tables from Jones *et al.*[58] and Jordan *et al.*[59]

TABLE 7 Risk-of-bias assessment form based on PROBAST (based on Wolff *et al.*[46])

Domain	Items
I. Participant selection	1a. Were appropriate data sources used, for example cohort, controlled trial or nested case–control study data?
	1b. Were all inclusions and exclusions of participants appropriate?
	1c. Were participants enrolled at a similar state of health, or were predictors considered to account for differences?
II. Predictors	2a. Were predictors defined and assessed in a similar way for all participants in the study?
	2b. Were predictor assessments made without knowledge of outcome data?
	2c. Are all predictors available at the time the model is intended to be used?
	2d. Were all relevant predictors analysed?
III. Outcome	3a. Was a prespecified outcome definition used?
	3b. Were predictors excluded from the outcome definition?
	3c. Was the outcome defined and determined in a similar way for all participants?
	3d. Was the outcome determined without knowledge of predictor information?
IV. Sample size and missing data	4a. Were there a reasonable number of outcome events?
	4b. Was the time interval between predictor assessment and outcome determination appropriate?
	4c. Were all enrolled participants included in the analysis?
	4d. Were participants with missing data handled appropriately?
V. Analysis	5a. Were non-binary predictors handled appropriately?
	5b. Was selection of predictors based on univariable analysis avoided?
	5c. Was model overfitting (optimism in model performance) accounted for, for example using bootstrapping or shrinkage techniques?
	5d. Were any complexities in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?
	5e. Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?
	5f. For the model or any simplified score, were relevant performance measures evaluated, for example calibration, discrimination, (re)classification and net benefit?
	5g. Was the model recalibrated or was it likely (based on the evidence presented, e.g. calibration plot) that recalibration was not needed?

Note
Partly reproduced from an early version of the PROBAST with permission from Dr Karel GM Moons, Utrecht University, 2020, www.probast.org (accessed 15 January 2020). The current PROBAST has been published previously.[48,49]

TABLE 8 Systematic review 2: reviews included

Review	Aim	Overlap with included studies in SR2
Schmidt-Hansen 2017[41]	To review the existing risk prediction tools for patients presenting in primary care with symptoms that may indicate lung cancer	Hamilton 2005,[51] Hippisley-Cox 2011,[52] Hippisley-Cox 2013,[53] Hippisley-Cox 2013[54] and Iyen-Omofoman 2013[21]
Elias 2017[50]	To validate published diagnostic models for their ability to safely reduce unnecessary endoscopy referrals in primary care patients suspected of significant colorectal disease	Fijten 1995,[55] Marshall 2011,[22] Muris 1995[56] and Nørrelund 1996[57]

Schmidt-Hansen *et al.*[41] concluded that, so far, none of the tools has been externally validated, yet there is a need to improve early diagnosis.

Elias *et al.*[50] aimed to identify and validate published diagnostic models to safely reduce unnecessary endoscopy referrals in CRC. A systematic review of the literature was undertaken and identified models were validated using a cross-sectional Dutch data set ($n = 810$). The definition of model used by Elias *et al.*[50] was very broad and included guidelines and weighted scores. Therefore, although Elias *et al.*[50] identified 18 models, only four are relevant to our review: Fijten *et al.*[55] and Marshall *et al.*[22] were previously identified from our searches, whereas Muris *et al.*[56] and Nørrelund *et al.*[57] are new inclusions. Because Elias *et al.*[50] attempted to validate the models they found, their validation of these four models is included in the results presented later in this chapter.

Prediction models

The 41 included studies (39 identified from the searches plus two identified from Elias *et al.*[50]) reported on 12 different prediction models (which are briefly summarised in the following paragraphs): (1) a RAT for 15 different cancer sites; (2) QCancer for six cancer sites, plus male and female versions for all cancers; (3) a clinical prediction rule for breast cancer;[60] (4) the BB equation for CRC;[22] (5) the Netherlands model for CRC;[55] (6) a machine-learning algorithm for CRC;[61] (7) a Danish model for CRC;[57] (8) a UK model for lung cancer;[21] (9) a UK model for pancreatic cancer;[62] (10) a European model for abdominal cancers;[63] (11) a UK model for paediatric cancers;[64] and (12) a Dutch model for multiple cancers.[56]

The RATs were designed to be used with patients presenting to primary care with 'low-risk-but-not-no-risk symptoms'.[65] Early versions of RATs were developed using case–control data from Devon, UK, as part of the CAPER studies.[39] Later models were derived using UK-wide primary care data – the Clinical Practice Research Datalink (CPRD) (formerly known as the General Practice Research Database),[66–73] and The Health Improvement Network (THIN) database.[74,75] So far, models for 14 separate cancer sites have been published (colorectal, oesophageal, lung, ovarian, kidney, bladder, pancreas, breast, uterine, brain, prostate, Hodgkin lymphoma, non-Hodgkin lymphoma and multiple myeloma), plus one model for metastatic cancer. The RATs are available as prints on common office objects (e.g. mouse mats) and are integrated into GP 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 *et al.*[50] used a Dutch data set to externally validate the colorectal version of RATs.

The QCancer series of models can be used in both symptomatic (diagnostic models) and asymptomatic (prognostic models) patients.[42] QCancer was developed in the QResearch database, a large database comprising > 12 million anonymised health records from 602 general practices throughout the UK, using the EMIS Health (Leeds, UK) computer system. Initially, several models were developed for each cancer type in symptomatic populations (colorectal, gastro-oesophageal, lung, renal, pancreatic and ovarian cancer). An updated approach incorporates multiple risk factors and symptoms into one model to predict cancer risk. Most of these models have been externally validated in UK-wide populations (e.g. THIN database[76]). QCancer is available as an online calculator (www.qcancer.org), which provides estimates of the absolute risk of any cancer, with a breakdown of type of cancer based on both risk factors such as age, sex 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.

McCowan *et al.*[60] developed a clinical prediction model for breast cancer using secondary care data on symptomatic patients at one hospital in Scotland. The authors validated the model using data from 202 patients with symptomatic breast problems attending 11 general practices in Scotland.

Marshall *et al.*[22] used data from the THIN data set (> 40,000 participants) to construct a model for CRC, known as the BB equation, which they validated using the CAPER data set. Data from 290 patients presenting to GPs in the Netherlands with rectal bleeding (from 1988 to 1990) were used by Fijten *et al.*[55] to develop a prediction model for CRC. Two studies validated this model: Hodder *et al.*[77] used secondary

care data from the UK, whereas Elias *et al.*[50] used a Dutch data set. Kop *et al.*[61] used a machine-learning algorithm to develop a prediction model for CRC using the electronic records of almost 220,000 patients from two general practices in the Netherlands. A Danish CRC model[57] has also been developed for use in primary care; this was externally validated by Elias *et al.*[50] using a Dutch data set.

Iyen-Omofoman *et al.*[21] developed a prediction model for lung cancer using data on > 130,000 participants in the THIN data set. Keane *et al.*[62] also used the THIN data set and developed two prediction models: one for pancreatic ductal adenocarcinoma and one for biliary tract cancers.

Dommett *et al.*[64] is the only model our searches identified that considers paediatric cancers. The authors used the CPRD to develop prediction models of bone and soft tissue, central nervous system and abdominal cancers, and leukaemia and lymphoma. The authors also developed a model to consider all paediatric cancers.

Muris *et al.*[56] developed a model using data from the Netherlands to predict multiple cancers, and Elias *et al.*[50] externally validated it.

Holtedahl *et al.*[63] report details of 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 in Norway, Denmark, Sweden, Scotland, Belgium and the Netherlands over a 10-day period, were used to develop the model. No validation of the model is reported.

Prediction model characteristics

Colorectal cancer was associated with the greatest number of models (six in total): (1) the BB equation,[22] (2) the Netherlands model,[55] (3) the machine-learning algorithm,[61,78,79] (4) the Danish model,[57] (5) QCancer[80] and (6) the RAT.[72,74,81] We identified three models for lung cancer (UK model,[21] QCancer[52] and RAT[51]) and three models for pancreatic cancer (UK model,[62] QCancer[32] and RAT[70]). Only versions of QCancer and RAT were found for gastro-oesophageal cancer,[71,82] ovarian cancer[83,84] and renal cancer.[66,85] There are two RATs for blood cancers: one for leukaemia[67] and one for myeloma.[68] For the other cancer sites, only one model for each was identified, and, apart from the breast cancer model,[60] the metastatic cancer model[86] and the abdominal model,[63] they are versions of RATs. Two versions of the QCancer model (one for females[53] and one for males[54]), a Dutch model[56] and the UK paediatric cancers model[64] were all developed to evaluate the risk prediction of multiple cancers.

The models are in various stages of development. A total of 18 models (or versions of models) have assessed only apparent performance, three models have been internally validated using a split-sampling technique[21,52–54] and one model was updated as a result of using a different data source.[74] Five of the QCancer versions,[32,80,82,84,85] one RAT version[81] and five of the other prediction models[22,55–57,60] have been externally validated, which is the highest level of evidence identified in this systematic review.

All but one of the models were developed in primary care settings. McCowan *et al.*[60] developed a clinical prediction rule model for breast cancer using secondary care data, but with the intention of the model being used in primary care. Only four models were developed outside the UK: those by Fijten *et al.*,[55] Kop *et al.*[61] and Muris *et al.*[56] were developed in the Netherlands, and the one by Nørrelund *et al.*[57] was developed in Denmark. For the models that were externally validated, most were validated in the country in which they were developed, except for the following: the validation[77] of the Netherlands CRC model[55] in a UK population, the validation of the Danish CRC model[57] in a Dutch population,[50] and the validation of the colorectal version of RATs (UK)[81] in a Dutch population.[50]

Table 9 provides a brief description of the models, their stages of development, the cancer sites covered and study designs. Owing to the heterogeneity in tools, cancer sites, outcomes measured and study design, a narrative review of the studies was conducted.

TABLE 9 Systematic review 2: summary of the prediction models, their stages of development, the cancer sites covered and the study designs

Cancer site and prediction model	Number and categories of descriptors	Stage of development	Study design	Country	Source
Bladder					
RAT	8; symptoms, medical history, test results	Apparent performance	Case–control	UK	Shephard 2012[69]
Blood					
RAT (leukaemia)	10 (chronic leukaemia); symptoms 13 (acute leukaemia); symptoms	Apparent performance	Case–control	UK	Shephard 2016[67]
RAT (myeloma)	16; symptoms, test results	Apparent performance	Case–control	UK	Shephard 2015[68]
Brain					
RAT	8; symptoms	Apparent performance	Case–control	UK	Hamilton 2007[87]
Breast					
Clinical prediction rule	5; patient demographics, symptoms	External validation	Prospective cohort	UK	McCowan 2011[60]
Colorectal					
BB equation	8; symptoms, test results	External validation	Retrospective case–control	UK	Marshall 2011[22]
		External validation	Prospective cohort	The Netherlands	Elias 2017[50]
The Netherlands model	3; symptoms, patient demographics	Apparent performance	Prospective cohort	The Netherlands	Fijten 1995[55]
		External validation	Prospective cohort	UK	Hodder 2005[77]
		External validation	Prospective cohort	The Netherlands	Elias 2017[50]
Machine learning algorithm	Numerous models are reported; patient demographics, symptoms, medical history, test results	Apparent performance	Case–control	The Netherlands	Kop 2015;[61] Kop 2016;[79] and Hoogendoorn 2015[78]
Danish model	2; patient demographics, symptoms	Apparent performance	Prospective cohort	Denmark	Nørrelund 1996[57]
		External validation	Prospective cohort	The Netherlands	Elias 2017[50]
QCancer	6 (females) and 7 (males); symptoms, medical history, test results	Internal validation II	Open prospective cohort	UK	Hippisley-Cox 2012[80]
		External validation	Prospective cohort	UK	Collins 2012[15]

continued

TABLE 9 Systematic review 2: summary of the prediction models, their stages of development, the cancer sites covered and the study designs (*continued*)

Cancer site and prediction model	Number and categories of descriptors	Stage of development	Study design	Country	Source
RAT	10; symptoms, test results	Apparent performance	Case–control	UK	Hamilton 2005[81]
		External validation	Prospective cohort	The Netherlands	Elias 2017[50]
RAT	8; symptoms, test results	Apparent performance	Case–control	UK	Hamilton 2009[74]
RAT (bowel)	10; symptoms, test results	Apparent performance	Case–control	UK	Stapley 2017[72]
Gastro-oesophageal					
QCancer	7 (females) and 6 (males); symptoms, test results, patient demographics	Internal validation II	Open prospective cohort	UK	Hippisley-Cox 2011[82]
		External validation	Retrospective cohort	UK	Collins 2013[88]
RAT	16; symptoms, test results	Apparent performance	Case–control	UK	Stapley 2013[71]
Lung					
UK (Iyen-Omofoman *et al.*[21]) model	15; patient demographics, symptoms	Internal validation II	Case–control	UK	Iyen-Omofoman 2013[21]
QCancer	9 (females) and 8 (males); symptoms, patient demographics, test results	Internal validation II	Open prospective cohort	UK	Hippisley-Cox 2011[52]
RAT	13; symptoms, patient demographics	Apparent performance	Case–control	UK	Hamilton 2005[51]
Ovarian					
QCancer	8; medical history, symptoms, test results	Internal validation II	Open prospective cohort	UK	Hippisley-Cox 2011[84]
		External validation	Retrospective cohort	UK	Collins 2013[89]
RAT	7; symptoms	Apparent performance	Case–control	UK	Hamilton 2009[83]
Pancreas					
QCancer	7 (females) and 8 (males); patient demographics, medical history, symptoms	Internal validation II	Open prospective cohort	UK	Hippisley-Cox 2012[32]
		External validation	Retrospective cohort	UK	Collins 2013[14]
RAT	9; medical history, symptoms	Apparent performance	Case–control	UK	Stapley 2012[70]
UK models for PDAC and BTC	13 (PDAC) and 9 (BTC); symptoms, medical history, test results	Apparent performance	Case–control	UK	Keane 2014[62]

Cancer site and prediction model	Number and categories of descriptors	Stage of development	Study design	Country	Source
Prostate					
RAT	9; symptom, test results	Apparent performance	Case–control	UK	Hamilton 2006[90]
Renal					
QCancer	7 (females) and 5 (males); medical history (females), patient demographics, symptoms, test results	Internal validation II	Open prospective cohort	UK	Hippisley-Cox 2012[85]
		External validation	Retrospective cohort	UK	Collins 2013[91]
RAT	15; symptoms, test results	Apparent performance	Case–control	UK	Shephard 2013[66]
Uterine					
RAT	9; symptoms, test results	Apparent performance	Case–control	UK	Walker 2013[73]
Metastatic					
RAT	7; symptoms, test results	Apparent performance	Case–control	UK	Hamilton 2015[86]
Multiple					
QCancer (female)	• 7 (uterine) • 10 (breast, blood) • 11 (ovarian, renal) • 12 (cervical) • 13 (colorectal, gastro-oesophageal) • 14 (pancreatic) • 15 (lung) • 22 (other cancers) • Medical history, symptoms, test results, patient demographics	Internal validation II	Open prospective cohort	UK	Hippisley-Cox 2013[53]
QCancer (male)	• 3 (testicular) • 8 (renal tract) • 12 (colorectal) • 13 (gastro-oesophageal) • 14 (prostate, blood) • 15 (pancreatic) • 17 (lung) • 20 (other cancers) • Medical history, symptoms, test results, patient demographics	Internal validation II	Open prospective cohort	UK	Hippisley-Cox 2013[54]

continued

TABLE 9 Systematic review 2: summary of the prediction models, their stages of development, the cancer sites covered and the study designs (*continued*)

Cancer site and prediction model	Number and categories of descriptors	Stage of development	Study design	Country	Source
UK paediatric model	• 4 (bone and soft tissue) • 7 (central nervous system, abdominal) • 12 (leukaemia/lymphoma, all cancers) • Symptoms	Apparent performance	Prospective case–control	UK	Dommett 2013[64]
Muris (the Netherlands) model	5; symptoms, patient demographics, test results	Apparent performance	Prospective cohort	The Netherlands	Muris 1995[56]
		External validation	Prospective cohort	The Netherlands	Elias 2017[50]
Abdominal model	4; symptoms, patient demographics	Apparent performance	Prospective cohort	Belgium, Denmark, the Netherlands Norway, Scotland and Sweden	Holtedahl 2018[63]

BTC, biliary tract cancer; PDAC, pancreatic ductal adenocarcinoma.

Critical appraisal

The assessment of risk of bias is summarised in *Table 10*, and is given in more detail in *Appendix 3* (see *Tables 41* and *42*). 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. Most of the included models were judged as having low risk of bias for participant selection and aspects of the outcome definition. For a number of studies, there was a high risk of bias surrounding aspects of the analysis. Much of this stemmed from methods reported for the selection of model variables based on univariate analyses, a lack of evaluation of calibration or discrimination, or no accounting for overfitting. For features of the predictors and participant flow, there was much uncertainty as to what had been done in the studies, and so the risk of bias of the findings could not be determined.

Discussion

Looking at symptom-based cancer diagnostic prediction models currently under development, we were able to identify 43 studies in total. The majority of these reported on various aspects of just two such modelling versions, QCancer and RAT, both developed in the UK. Most of the reported work also seemed to be located in Europe, with two of the models developed in the Netherlands.

The majority of the models included in this review were developed only with the sample used to derive the model. With the exception of the RAT (colorectal) model, there were no reports to suggest that models were being updated based on new available data. Some models were validated using split-sample techniques. A number of models, in particular the ones developed under the QCancer name, were externally validated in independent studies.

The two main models highlight important knowledge gaps; the development of the QCancer models was based on higher-quality data (cohort data) than that of the RATs, and were mostly externally validated, but lack impact assessment. By contrast, the RAT series has more evidence of impact on practice, but was developed from case–control studies and has limited external validation.

Our systematic review was limited to the inclusion of diagnostic prediction models; however, the search strategy also highlighted a number of studies concerned with the development of symptom-based decision tools that were not based on prediction models, such as scoring systems, as well as the development of diagnostic prediction models for secondary care. These are listed in *Appendix 3* (see *Tables 43* and *44*) as a resource for further research.

Evidence suggested that there has been a great deal of external validation work for QCancer, whereas we found only one attempt at external validation of RATs. Ideally, this is an area for further development of the RATs and the four other models that have not yet been externally validated.

Conclusion

To our knowledge, this is the first systematic review of diagnostic prediction models for use in primary care to aid cancer diagnosis. We have identified models that have the potential to be developed into tools and be used by GPs. However, there are gaps in the literature that we have identified.

Currently, most research on developing symptom-based cancer risk prediction models is concentrated in Europe and, in particular, the UK. QCancer and RATs are the dominant prediction models. Although there has been a great deal of external validation work done for QCancer, only one study was identified that reports the external validation of RATs.

TABLE 10 Systematic review 2: risk-of-bias assessment of the included models

Model (first author of first version)	Stage of development covered	Risk-of-bias domain				
		I. Participant selection[a]	II. Predictors[a]	III. Outcome[a]	IV. Sample size and participant flow[a]	V. Analysis[a]
RAT (Hamilton) series of models for multiple sites[51,66–74,81,83,86,87,90]	Apparent performance	✓	?	✓	?	✗
	External validation (colorectal)[50]	✓	✓	✓	?	?
QCancer (Hippisley-Cox) series of models for multiple sites and populations (female/male)[14,15,32,52–54,80,82,84,85,88,89,91]	Internal validation II	✓	✓	✓	✓	✓
	External validation (not for lung, 'male', or 'female')	✓	✓	✓	✓	✓
Clinical prediction rule (McCowan 2011[60]) for breast cancer	External validation (developed in secondary care for use in primary care)	✓	✓	✓	✗	✗
BB (Marshall)[22] model for CRC	External validation	✓	?	✓	?	✓
	External validation (Elias 2017[50])	✓	✓	✓	?	?
The Netherlands' (Fitjen 1995[55]) model for CRC	Apparent performance	✗	✓	✓	?	✗
	External validation (Hodder 2005[77])	✗	?	✗	✓	?
	External validation (Elias 2017[50])	✓	✓	✓	?	?
The Netherlands' (Kop)[79] 'machine learning' for CRC	Apparent performance	✓	?	✓	?	?
Danish (Norrelund 1996[57]) model for CRC	Apparent performance	✓	?	✓	?	✗
	External validation (Elias 2017[50])	✓	✓	✓	?	?
The Netherlands' (Muris 1995[56]) model for CRC	Apparent performance	?	✓	✓	?	✗
	External validation (Elias 2017[50])	✓	✓	✓	?	?
UK (Iyen-Omofoman 2013[21]) model for lung cancer	Internal validation II	✓	?	✓	?	✗
UK (Keane 2014[62]) model for pancreatic cancer	Apparent performance	✓	✓	✓	?	✓
UK (Dommett 2013[64]) model for paediatric cancer	Apparent performance	✓	✓	✓	✓	?
Prediction model for abdominal cancers (Holtedahl 2018[63])	Apparent performance	?	✓	?	✗	?

✓, low risk of bias; ✗, high risk of bias; ?, unclear risk of bias.
a Multiple ordered by stage of development if different.

Chapter 5 Updated review

Introduction

An update of a previous systematic review by Neal *et al.*[25] (published in 2015) was undertaken to examine the association between different durations of time from first symptom to diagnosis or treatment and clinical outcomes across all major cancers for symptomatic presentations, to investigate whether or not a more timely cancer diagnosis is associated with more favourable outcomes. It had been anticipated that there would be very little evidence on whether or not diagnostic tools have led to increased or quicker cancer diagnoses and, ultimately, impacts on patient survival. This updated review would therefore provide important information to inform the proposed decision model of a cancer diagnostic pathway for assessing the impact of using diagnostic tools.

The cancer diagnostic pathway is complex, incorporating multiple key time points and intervals from the patient first experiencing symptoms to receiving a definitive diagnosis or treatment (*Figure 2*). An initiative to improve the design and reporting of early-cancer diagnosis research included a review of existing

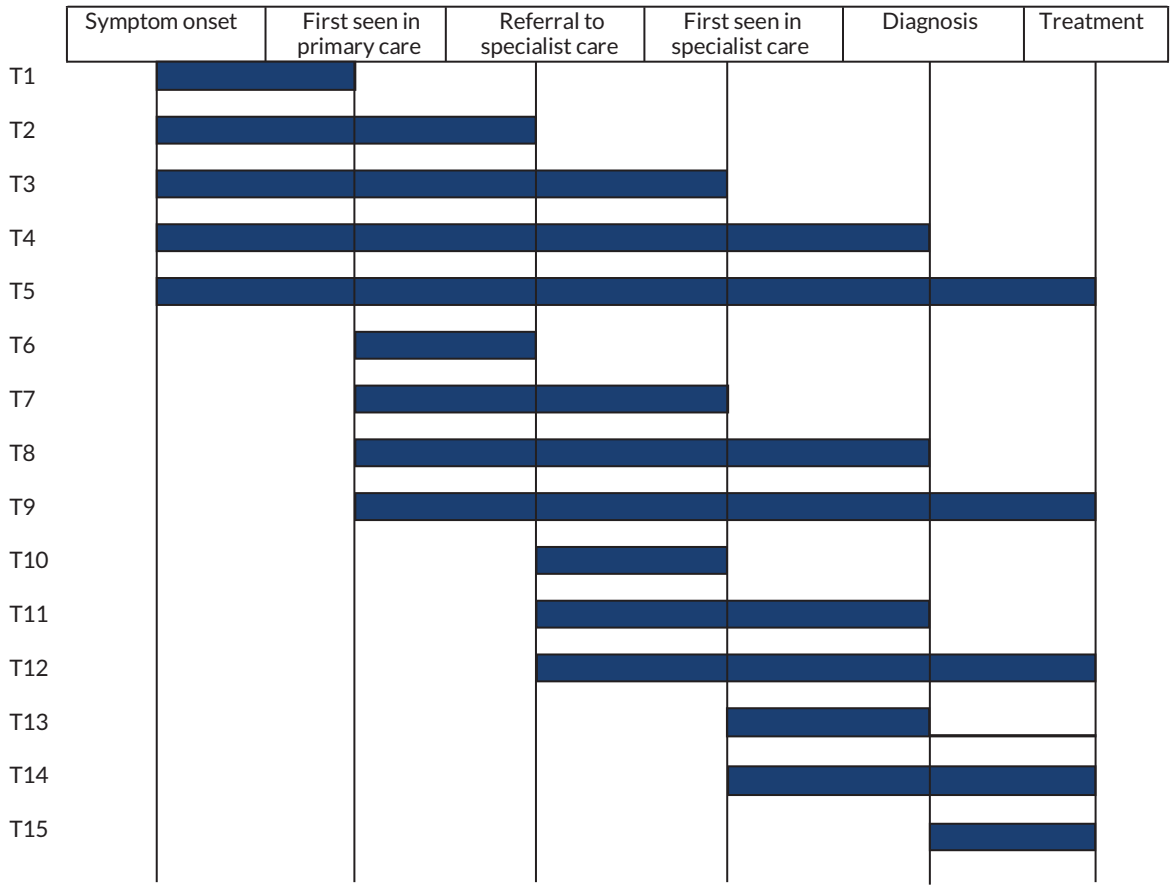


FIGURE 2 Illustration of the different interval types. T1, time from symptom onset to first seen in primary care ('patient interval'); T2, time from symptom onset to referral to specialist care; T3, time from symptom onset to first seen in specialist care; T4, time from symptom onset to diagnosis; T5, time from symptom onset to treatment; T6, time from first seen in primary care to referral to specialist care ('referral interval'); T7, time from first seen in primary care to first seen in specialist care; T8, time from first seen in primary care to diagnosis ('diagnostic interval'); T9, time from first seen in primary care to treatment; T10, time from referral to specialist care to first seen in specialist care; T11, time from referral to specialist care to diagnosis; T12, time from referral to specialist care to treatment; T13, time from first seen in specialist care to diagnosis; T14, time from first seen in specialist care to treatment; T15, time from diagnosis to treatment (treatment interval).

instruments used to measure time points and intervals was undertaken. The project also incorporated a consensus conference approach combined with nominal group techniques to develop a series of recommendations for definitions and methodological approaches, which culminated in the Aarhus consensus statement, published in 2012.[92] The Aarhus checklist is a resource for early-cancer diagnosis research that aims to promote greater precision and transparency in both definitions and methods used.[92]

The broader evaluation of how reductions in time to diagnosis or treatment can affect patient outcomes with any cancer not only allows the consideration of the probable impact of using diagnostic tools in these cancers, but also provides a greater evidence base on the probable impact that could be expected from reducing time to diagnosis and/or treatment in general. A more in-depth evaluation of CRC studies was undertaken in this updated review, as this was the focus of our decision-analytic model. This in-depth evaluation focused on the methodology used in those studies. The decision-analytic model focused on the use of diagnostic tools in an adult population; therefore, unlike the previous review of diagnostic intervals by Neal *et al.*,[25] we did not consider childhood cancers as part of the updated review. The previous review and this updated review considered any type of interval along this pathway, which were grouped according to accepted definitions, relating to the patient, primary care, secondary care or combinations of any of those categories.[92]

Aim

The aim of this review was to update a previous systematic review[25] assessing the evidence linking the durations of different intervals in the diagnostic process to clinical outcomes. A more in-depth evaluation of CRC studies, focusing on methods, was conducted to identify studies that are likely to provide the most valid estimate of the impact of diagnostic intervals on patient outcomes for informing the decision-analytic model.

Methods

Identification of studies

Search strategy

Searches conducted for the previous review and developing the methods for the current update

The previous review[25] was conducted in two phases. The original review was conducted in 2008–10, and then a subsequent review was conducted in 2013–14; both phases were published as a single review in 2015. The first phase did not include breast cancer or CRC, because of existing systematic reviews (e.g. Richards *et al.*;[93] Ramos *et al.*;[94] Ramos *et al.*;[95] and Thompson *et al.*[96]). However, these cancer sites were included in the subsequent phase (in 2013–14).

The literature searches for the previous review, which covered multiple databases, were first conducted during Phase I in 2010 (covering the literature from inception of the databases to February 2010) and then during Phase II in 2013 (covering the literature from February 2010 to November 2013). Both the original and subsequent searches (Neal *et al.*;[25] Phases I and II) yielded a very high number of references to screen. Our estimate from scoping searches was that an update of the Neal *et al.*[25] review from 2013 to 2018 would yield some 25,000 studies to screen. We therefore adopted a pragmatic alternative method to update this review, using forward citation-chasing.

Search strategy for the current updated review

The forward citation searches for the current update were conducted in August 2017 and updated in February 2018. Forward citations of the original list of 177 studies included in Neal *et al.*[25] (from both the 2010 and the 2013 searches) were chased using Scopus and the Web of Science. The search

results were exported to EndNote X7 and de-duplicated using automatic and manual checking, yielding 2769 studies for screening. The full text of included studies identified from screening were also forward citation-chased (second-order chasing) in the same way until the investigations were exhausted and all included studies had been citation-chased.

It is uncertain as to what is the best method for updating systematic reviews.[97] The value of alternative methods of searching is acknowledged, but more research is needed on the clinical effectiveness of different techniques.[98] To test the thoroughness of our novel approach, we sought forward citations of all the full-text references in the 2010 searches and compared the results with those of the full updated review carried out in 2013. The citation-chasing approach found 34 of the 71 full-text papers identified in the 2013 updated review. However, the 2013 review also incorporated some revisions to the inclusion criteria to make the review more focused, making results less directly comparable.

Selection of studies

Two independent reviewers (RL and BG) screened the titles and abstracts of all records identified by the searches for relevance, and then assessed the potentially relevant records subsequently retrieved as full texts for inclusion. Disagreements were resolved by discussion or, if necessary, taken to a third reviewer (JP).

Eligibility criteria

Studies were included if they fulfilled the following criteria:

- They primarily set out to determine the association of at least one time interval to diagnosis or treatment with patient outcomes.
- They included symptomatic adult patients with primary cancers (excluding screening- and biomarker-detected cancers).
- They investigated at least one diagnostic interval (patient, primary care, or a combination), which could be assessed against accepted definitions (Aarhus statement[92]). Studies that investigated the impact of time from diagnosis to treatment were included only if they also reported data on pre-diagnostic time. (The diagnostic pathway begins at the time of symptom onset and ends at the point of definitive diagnosis or treatment, whereas the interval between diagnosis and treatment falls within the treatment pathway and generally relates to secondary care only.)
- They reported data on survival, morbidity, or stage at diagnosis.
- They were available as a full-text paper in English.

Studies comparing two or more diagnostic or referral pathways (e.g. NICE 2WW referrals vs. non-urgent referrals) were included only if (overall) group-level numeric values for each interval (see *Figure 2*) were reported, that is the study also assessed the impact of different intervals on patient outcomes, irrespective of the referral pathway.

Data extraction

Data extraction was undertaken by one reviewer (RL) and checked by another (JP). For studies of CRC, this included data on a study's aims; design; population; location; setting; number of participants sampled and, subsequently, recruited and analysed; definitions of time duration; data collection methods; outcome measures used; and the authors' conclusions. The main results were also extracted, including the methods used for assessing the association between interval and outcomes, statistical significance, CIs and any subgroup analysis. A more condensed summary was extracted for studies of other cancer sites.

Critical appraisal

The methodological quality of each CRC study included was assessed using the same bias assessment tool used in the previous review[25] (see *Appendix 4*, *Box 1*). Bias assessment was undertaken by two independent reviewers (RL and JP); there were no disagreements that needed be taken to a third person.

Waiting-time paradox

The risk-of-bias assessment also included identifying studies that addressed the 'waiting-time paradox', as they are likely to be of better analytical quality.

The observed association between a short diagnostic interval and poor outcomes has been referred to as the 'waiting-time paradox'. Two underlying theories have been proposed to explain this effect. One widely held belief is confounding by severity of disease, whereby high-risk precursors, such as aggressiveness of the tumour, act as unmeasured confounders that mask the effect of the exposure.[99] The underlying assumption here is that rapidly growing or more aggressive tumours are more likely to present with alarm symptoms that will make the patient or doctor think of cancer, whereas slow-growing or less well-differentiated tumours result in vague initial symptoms that are difficult to detect.[99] To help mitigate this bias, information is required on factors associated with the aggressiveness of the tumour, such as histology, grade, stage, tumour volume, genetic factors. The alternative assumption is that the paradoxical findings reflect confounding by indication, caused by differentiated clinical triage.[99] The assumption here is that GPs expedite patients presenting with high-risk symptoms, especially if they look ill, and at the same time are more reluctant to refer healthy-looking people with low-risk symptoms.[99] To help mitigate this bias, information is required on what triggers the GP to either refer immediately or adopt a watchful waiting approach.[99]

In the previous systematic review, studies that addressed the waiting-time paradox were defined as:

> *. . . articles that undertake an analysis or sub-analysis that specifically includes or excludes patients who are either diagnosed very quickly (e.g., within 4–8 weeks, although this will vary between cancers), or have very poor outcomes (e.g., deaths within a short time after diagnosis, e.g., within 4–8 weeks).*
>
> *Neal* et al.[25]

Papers that simply reported that the waiting-time paradox may have confounded their data were not classified as having addressed this potential source of bias. In the current updated review, we recorded whether or not studies had performed adjusted analyses for aspects such as emergency presentation, severity of presenting symptoms, tumour grade or aggressiveness. This is discussed in more detail in *Colorectal cancer*.

Data synthesis

A narrative synthesis was undertaken, which followed the same approach as that used in the previous review.[25] Overall findings were presented in a table format, with studies grouped by cancer location. Information reported included type of diagnostic interval, and whether the analysis considered patient outcomes, such as survival or stage of disease, or other outcomes. A key output was to identify the association between reduced diagnostic interval and improved patient outcome. This information fed directly into the decision-analytic model and permitted an exploration of the variation in findings from different studies. The previous review[25] identified a large degree of variability between included studies.

In the table of overall findings, studies that reported 'positive' associations (i.e. there was evidence of shorter intervals being associated with more favourable outcomes) were presented first. Studies that reported no associations were reported next. Studies that reported 'negative' associations (i.e. there was evidence of shorter intervals being associated with less favourable outcomes) were reported at the end of the table. Within each grouping, studies were ordered by the patient outcomes: survival or mortality first (for simplicity, both are referred to as survival in the table), followed by stage and then 'other' outcomes.

Figure 2 summarises the definitions that were used to categorise the different time (T) durations evaluated by included studies. Each duration was categorised as one of 15 intervals.

Data synthesis

The results of the current updated review are first presented for all cancers using the same table format as in the previous review.[25] They are followed by more in-depth evaluation for CRC. No meta-analyses were planned or conducted as a result of the (anticipated) heterogeneity: the review of CRC therefore focuses on identifying studies that are likely to provide the best estimate of the impact of diagnostic intervals on patient outcomes for informing the decision-analytic model.

The CRC review incorporates studies identified during our updated review and better-quality studies identified during the previous review. We focus only on more recent or better-quality studies as they are considered to provide more valuable information for our decision model for the following reasons:

- They are more likely to account for the *non-linear association* between time to diagnosis and stage or survival.
- They are more likely to account for the *waiting-time paradox*, whereby patients with more advanced or aggressive disease are inclined to present earlier and have definitive symptoms, or are prone to be referred quickly and prioritised for treatment.
- They use *clearer definitions of diagnostic intervals*, following the publication of the Aarhus consensus statement in 2012.[92]

Results

Studies identified

Our searches identified 1810 references after de-duplication, of which 104 were considered relevant and retrieved as full-text studies. Thirty-five studies were identified that met the inclusion criteria (*Figure 3*). This includes one study for which the interlibrary loan was still outstanding at the time of the report; therefore, the data extraction is based on an abstract. This study[122] included patients with well-differentiated neuroendocrine tumours.

A summary of the included studies is provided in *Appendix 4* (see *Table 51*). Two studies considered multiple cancer sites, whereas the remaining studies evaluated a sole cancer: bladder ($n = 1$), breast ($n = 4$), colorectal ($n = 8$), lung ($n = 5$), lymphoma ($n = 1$), myeloma ($n = 1$), neuroendocrine ($n = 1$), oral ($n = 2$), ovarian ($n = 2$), pancreatic ($n = 2$), penile ($n = 1$), prostate ($n = 1$), sarcoma ($n = 3$) and testicular ($n = 1$). The studies were from a wide range of countries. Eight studies were undertaken in the UK, five in Canada and four in Spain; two studies were undertaken in each of the following countries: the USA, the Netherlands, Mexico and Japan; and one study was carried out in each of the following countries: China, France, Islamic Republic of Iran, Israel, Italy, Montenegro, Poland, Rwanda, Sri Lanka and Uganda.

Findings of the previous review

The previous review[25] identified a large number of relevant studies ($n = 209$), with only a small number considered to be of high quality.[25] There was substantial variation in the type of intervals evaluated and the findings within and between cancer types. Compared with other cancer types, studies of colorectal, breast, head and neck, and testicular cancers and melanoma suggested that shorter time intervals were associated with improved patient outcomes. However, for each of these cancer types, there were also studies reporting no association between time interval and patient outcome. Some of the included studies also showed that shorter intervals were associated with decreased survival. This important heterogeneity is likely to have been caused by confounding, as discussed in *Waiting-time paradox*. A summary table of the previous review findings is provided in *Appendix 4* (see *Table 58*).

Findings of the updated review

A summary of the findings of the 35 new studies included in the updated review is presented in *Table 11*, using the same format as used in the original review. Some studies assessed the impact of longer intervals or perceived delay on patient outcomes, which we converted to represent the opposite effect associated

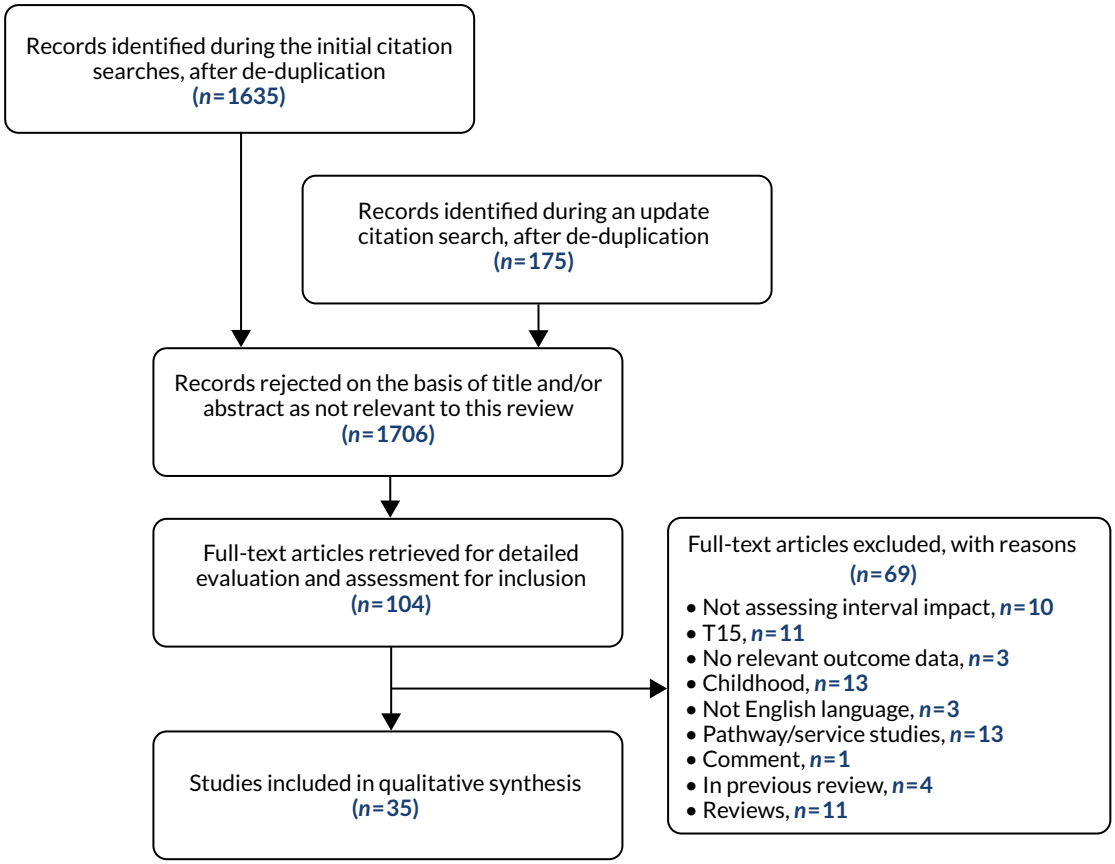


FIGURE 3 Updated review: the PRISMA flow diagram.

TABLE 11 Updated review: summary results from narrative synthesis, by cancer

Interval and outcome measure[a]	Time	Study		
		Positive association	No association	Negative association
Breast cancer				
Survival				
Patient interval	T1		Ángeles-Llerenas 2016[100]	
Symptom onset to treatment	T5		Ángeles-Llerenas 2016[100]	
Diagnostic interval	T8		[b]Redaniel 2015[101]	
Primary care to treatment	T9			Murchie 2015[102]
Treatment interval	T15		Ángeles-Llerenas 2016[100]	
Stage				
Patient interval	T1	Pace 2015[103]		
Symptom onset to specialist care	T3	Unger-Saldaña 2015[104]		
Diagnostic interval	T8	Pace 2015[103]		
Seen in primary care to treatment	T9			Murchie 2015[102]
From specialist care to treatment	T14	Unger-Saldaña 2015[104]		

TABLE 11 Updated review: summary results from narrative synthesis, by cancer (*continued*)

Interval and outcome measure[a]	Time	Study		
		Positive association	No association	Negative association
Bladder cancer				
Survival				
Symptom onset to referral	T2		Bryan 2015[105]	
Symptom onset to treatment	T5		Bryan 2015[105]	
Referral to first seen by specialist	T10		Bryan 2015[105]	
Referral to treatment	T12		Bryan 2015[105]	
Specialist care to treatment	T14		Bryan 2015[105]	
Stage				
Symptom onset to referral	T2		Bryan 2015[105]	
Symptom onset to treatment	T5		Bryan 2015[105]	
Referral to seen by specialist	T10		Bryan 2015[105]	
Referral to treatment	T12		Bryan 2015[105]	
Specialist care to treatment	T14		Bryan 2015[105]	
Other (tumour size)				
Referral to treatment	T12			Bryan 2015[105]
Specialist care to treatment	T14			Bryan 2015[105]
CRC				
Survival				
Primary care to treatment	T9		Helewa 2013;[106] Murchie 2014[107]	
Diagnostic interval	T8		[b]Redaniel 2015;[101] Dregan 2013[108]	
Symptom onset to diagnosis	T4		Pita-Fernández 2016[109]	
Referral to treatment	T12		Aslam 2017[110]	Patel 2018[111]
Stage				
Patient interval	T1		Chen 2017 (≥ 50 years)[112]	Chen 2017 (< 50 years)[112]
Symptom onset to diagnosis	T4		Chen 2017 (≥ 50 years);[112] Leiva 2017[113]	Chen 2017 (< 50 years)[112]
Diagnostic interval	T8		Chen 2017;[112] Leiva 2017 (GP records)[113]	Leiva 2017 (hospital records)[113]
Primary care to treatment	T9		Murchie 2014[107]	
Referral to treatment	T12			Patel 2018[111]
Other (metastases or nodal involvement)				
Referral to diagnosis	T11		Janssen 2016[114]	

continued

TABLE 11 Updated review: summary results from narrative synthesis, by cancer (*continued*)

Interval and outcome measure[a]	Time	Study		
		Positive association	No association	Negative association
Gastro-oesophageal cancer				
Survival				
Diagnostic interval	T8		[b]Dregan 2013[108]	
Lung cancer				
Survival				
Patient interval	T1		Radzikowska 2013 (SCLC);[115] Živković 2014[116]	
Symptom onset to specialist care	T3		Gonzalez-Barcala 2014[117]	
Symptom onset to diagnosis	T4		Živković 2014[116]	
Diagnostic interval	T8		[b]Dregan 2013[108]	[b]Radzikowska 2013 (SCLC);[115] Redaniel 2015[101]
Referral to diagnosis	T11		Gonzalez-Barcala 2014[117]	
Specialist care to treatment	T14			Gonzalez-Barcala 2014[117]
Treatment interval	T15			Gonzalez-Barcala 2014[117]
Stage				
Specialist care to diagnosis	T13		Gildea 2017[118]	Kim 2016 (NSCLC)[119]
Specialist care to treatment	T14			Kim 2016 (NSCLC)[119]
Treatment interval	T15			Kim 2016 (NSCLC)[119]
Lymphoma				
Survival				
Primary care to specialist	T7		Nikonova 2015[120]	
Specialist care to treatment	T14		Nikonova 2015[120]	
Myeloma				
Survival				
Symptom onset to diagnosis	T4		Goldschmidt 2016[121]	
Stage				
Symptom onset to diagnosis	T4		Goldschmidt 2016[121]	
Neuroendocrine tumours				
Survival				
Symptom onset to diagnosis	T4		Keizer 2016[122]	
Stage				
Symptom onset to diagnosis	T4		Keizer 2016[122]	

TABLE 11 Updated review: summary results from narrative synthesis, by cancer (*continued*)

Interval and outcome measure[a]	Time	Study		
		Positive association	No association	Negative association
Oral cancer				
Stage				
Patient interval	T1	Alahapperuma 2017[123]		
Referral to seen by specialist	T10		Alahapperuma 2017[123]	
Symptom onset to treatment	T5	Esmaelbeigi 2014[124]		
Ovarian cancer				
Stage				
Patient interval	T1		Lim 2016[125]	
Diagnostic interval	T8	Altman 2017[126]	Lim 2016[125]	
Pancreatic cancer				
Survival				
Symptom onset to diagnosis	T4	Gobbi 2013[127]		
Symptom onset to treatment	T5		Jooste 2016[128]	
Penile cancer				
Survival				
Patient interval	T1	Gao 2016[129]		
Stage				
Patient interval	T1	Gao 2016[129]		
Other (tumour size; metastases; lymph node involvement)				
Patient interval (tumour size)	T1	Gao 2016[129]		
Patient interval (metastases)	T1	Gao 2016[129]		
Patient interval (lymph node involvement)	T1	Gao 2016[129]		
Prostate cancer				
Stage				
Diagnostic interval	T8		Bonfill 2015[130]	[b]Redaniel 2015[101]
Treatment interval	T15		Bonfill 2015[130]	
Sarcoma				
Survival				
Symptom onset to specialist care	T3			Urakawa 2015[131]
Symptom onset to diagnosis	T4		Goedhart 2016[132]	
Referral to diagnosis	T11		Goedhart 2016[132]	

continued

TABLE 11 Updated review: summary results from narrative synthesis, by cancer (*continued*)

Interval and outcome measure[a]	Time	Study		
		Positive association	No association	Negative association
Stage				
Patient interval	T1	De Boer 2014[133]		
Other (metastases)				
Symptom onset to diagnosis	T4		Goedhart 2016[132]	
Testicular cancer *Survival*				
Patient interval	T1	Kobayashi 2014[134]		
Stage				
Patient interval	T1		Kobayashi 2014[134]	
Other (tumour size)				
Patient interval	T1	Kobayashi 2014[134]		
Urinary tract cancer *Survival*				
Diagnostic interval	T8	[b]Dregan 2013[108]		

NSCLC, non-small-cell lung cancer; SCLC, small-cell lung cancer.
a A positive outcome measure corresponds to better survival outcomes (or reduced mortality) for 'survival'; having a less advanced stage at diagnosis for 'stage'; and having a smaller tumour, no metastases, or no nodal involvement for 'other'.
b Study considered multiple cancer sites.
Notes
'Positive association' = shorter intervals associated with statistically significant more favourable outcomes.
'No association' = findings not statistically significant.
'Negative association' = shorter intervals associated with statistically significant unfavourable outcomes.
(Favourable outcomes = improved survival, earlier stage at diagnosis, or a less extensive tumour.)

with shorter intervals. The original interpretation of the results and how each was converted to show the corresponding impact of short intervals are presented in *Appendix 4* (see *Table 53*). Some of the studies that evaluated the impact of an interval (e.g. T1) as a categorical variable (e.g. 1–2, > 2–3 and > 3–4 months) reported statistically significant findings for some intervals and not others, the results of which are also shown in *Appendix 4* (see *Table 53*).

A larger number of studies identified during the updated review reported non-statistically significant findings or negative association between shorter intervals and patient outcomes. The potential of a 'U'-shaped association between time intervals and outcomes is explored in more detail in *Colorectal cancer*.

Colorectal cancer

The more in-depth summary for CRC presented here focuses on the 10 studies identified in our updated review (see *Table 11* for overall findings; this included two studies that considered multiple cancer sites) and four good-quality studies (published between 2011 and 2013) identified in the previous review that addressed the waiting-time paradox. The broader findings presented in the previous review,[25] and other previous systematic reviews by Ramos *et al.*[94,95] and Thompson *et al.*,[96] are considered to provide a good summary of the wider (prior) evidence base.

A more in-depth data extraction was conducted for the 14 included CRC studies. The results are provided in *Appendix 4*, presented over three tables summarising each study's characteristics (see *Table 55*), critical appraisal (see *Table 56*) and results (see *Table 57*). These results tables include a summary of both the magnitude and the direction (and statistical significance) of the associations between intervals and patient outcomes for each included study.

The following section presents the overall findings for the CRC studies, and an evaluation of the included studies that was conducted to aid the selection of specific studies to inform the decision model.

The overall findings

Five studies were undertaken in the UK, three in Denmark and two in each of the USA, Canada and Spain. Six studies were based in primary care, six in secondary care (including specialist centres) and two in both primary and secondary care. Four of the studies undertaken in secondary care were of single sites (hospitals). The three studies from Demark were by the same authors (i.e. Tørring *et al.*[99,135,136]) and included the same or overlapping study populations. Tørring *et al.*[136] focused solely on CRC, included 268 participants and data on estimated 3-year survival. Tørring *et al.*[135] analysed data from three population-based studies, one of which was Tørring *et al.*[136] Tørring *et al.*[99] evaluated the impact of diagnostic intervals in five common cancers, including CRC. The colorectal cohort for this study appears to be identical to that of Tørring *et al.*[136] ($n = 268$), but the analysis is based on 5-year survival. As our review is based on a narrative synthesis, all three studies[99,135,136] are included and described below.

The only interval to be evaluated by more than two studies was T8, which is the 'diagnostic interval' (time from first seen in primary care to diagnosis). The impact of the diagnostic interval on survival was evaluated by six studies,[99,101,108,135–137] one of which[137] evaluated colon and rectal cancers separately, and another of which[136] stratified analysis by whether the presenting symptoms were considered by the GP as alarm and/or serious versus vague. Four studies[99,135–137] found that longer intervals were associated with statistically significant poorer survival outcomes: one[137] in colon cancer only, and one[136] for alarm and/or serious symptoms only. The overall effect sizes that were reported by the studies (see *Appendix 4*, *Table 57*) were small, except for Tørring *et al.*,[135] which reported a medium size effect (odds ratio of 4.74) for the association between a very short interval and survival (Tørring *et al.*[135,136] also reported wide CIs). The diagnostic interval was analysed as both a categorical and a continuous variable in all three studies by Tørring *et al.*,[99,135,136] whereas it was evaluated as a dichotomised variable by Pruitt *et al.*[137] The diagnostic interval, when analysed as multiple categorical variables, was divided into three groups, based on the interval quartiles. The middle-interval group, which included both the second and the third quartiles combined, was used as the reference. The findings of the categorical analysis were statistically significant for the very-short-interval group in both Tørring *et al.*[136] (alarm or serious symptoms only) and Tørring *et al.*,[99] whereas Tørring *et al.*[135] reported significant results for both the very short- and the long-interval groups (see *Appendix 4*, *Table 57*). Tørring *et al.*[135] specifically aimed to test the theory of a U-shaped association between the diagnostic interval and mortality after diagnosis using data from three population-based studies.[81,136,138] The analysis of the combined data, from all three studies,[81,136,138] and the analysis of the data from the study based on GP data[138] showed that both the shorter- and the longer-interval groups were associated with a statistically significant increased risk of mortality at 5 years.[135] In all three studies by Tørring *et al.*,[99,135,136] the analysis and the graphical display of the diagnostic interval assessed as a continuous variable had a convex (U-shaped) association with mortality (these trends were statistically significant in Tørring *et al.*[99,135]). In other words, patients with very short or very long intervals had higher mortality than the rest. However, for patients presenting with vague symptoms,[136] there was a trend towards a concave association between diagnostic interval and mortality, but the results were not statistically significant and lacked power.

Two studies[112,113] assessed the impact of the diagnostic interval (T8) on stage at diagnosis. Leiva *et al.*[113] measured the interval using two separate data sources (GP and hospital records); Chen *et al.*[112] reported subgroup analysis according to age at diagnosis (< 50 and ≥ 50 years). Leiva *et al.*[113] found that shorter intervals were associated with more advanced cancer stage at diagnosis, based on data

from the hospital records only. Chen *et al.*[112] and Leiva *et al.*[113] also assessed the association between other intervals and stage at diagnosis, including duration between symptom onset and diagnosis (T4) (both studies) and the 'patient interval' (T1) (Chen *et al.*[112] only). Among younger patients (aged < 50 years), those with stages III or IV cancer had statistically significantly shorter durations for both T4 and T1 than those with stages I or II disease. One study[109] evaluated the association between T4 and survival, and found that longer symptoms-to-diagnosis intervals were not associated with poorer survival (or advanced stage) in CRC patients (see *Appendix 4*, *Table 57*).

Two further interval types were each investigated by two studies: T9[106,107] and T12.[110,111] Only Patel *et al.*[111] found one of these intervals (T12) to be associated with statistically significant findings: long-term survival and the proportion of patients with early-stage cancer were greatest in patients who did not receive treatment within the recommended 62 days (see *Appendix 4*, *Table 57*).

The intervals T11 (between referral to a specialist care and diagnosis)[114] and T15 (between definitive diagnosis and treatment)[137] were both evaluated by a single study. Janssen *et al.*[114] did not find shorter intervals to be associated with an increased likelihood of having distant metastases or a node-positive tumour. Pruitt *et al.*[137] found that shorter treatment intervals were associated with better survival in colon, but not rectal, cancer.

Summary of the methods used

Non-linear association

Tørring *et al.*[99,135,136] showed that the association between diagnostic interval and patient outcomes for CRC is U-shaped, with both very short and long intervals being associated with poor outcomes.[99,135,136] Five included studies used spline regression analysis to account for this non-linear relationship.[99,107,109,135,136] In the Murchie *et al.*[107] study, the spline curves for the unadjusted analyses showed a statistically significant non-linear association between time from being seen in primary care and treatment (T9, which the authors refer to as 'provider delay'[107]) and both stage and mortality, but these were no longer present after adjusting for confounders. Pita-Fernández *et al.*[109] presented spline regression analyses for colon and rectal cancers, which showed very short symptom-to-diagnosis (T4) intervals to be associated with higher mortality in those with rectal tumours. In colon cancer, no significant relationship was found between interval and survival. In addition to analysing the interval as a continuous variable, the study computed Kaplan–Meier curves for each interval quartile. The Cox regression model, adjusting for age and sex, showed that patients with cancer of the rectum in the first interval quartile had statistically significant lower survival than the rest, but this was no longer significant when controlling for stage; no significant difference was found for colon cancer.

Waiting-time paradox

Six studies[99,101,107,108,135,136] collected data on patients seen in primary care. Four studies[99,101,136,137] conducted a stratified analysis according to the GP's perceived seriousness of the presenting symptoms. Two studies[107,113] included the perceived seriousness (vague, serious, alarm) of the presenting symptom,[113] and/or specific (high-risk) symptoms[107,113] as covariates in their multivariate analyses. One study,[108] which evaluated the time intervals between specific alarm symptoms and subsequent cancer diagnosis, also compared the survival of patients with alarm symptoms with that of patients without alarm symptoms. One study[137] stratified diagnostic delay models by the four most common presenting symptom types as part of their sensitivity analysis, but did not present the results, noting only that these did not substantially change the findings.

A number of studies controlled for the confounding effect of tumour grade or aggressiveness. Potential important confounding factors include tumour stage, tumour grade, degree of differentiation, type of hospital admission (emergency or elective), symptoms and signs. Five studies[101,106,109,113,137] used multivariate analysis that accounted for one or more of these factors. Six studies also conducted stratified analysis according to stage,[137] emergency admission[106,110,111] or presenting symptoms (either type of or perceived seriousness).[99,101,137] Emergency admission was also considered as an exclusion criterion in two studies.[101,114]

However, some argue that stage at diagnosis can be conceived as a mediator or intermediate factor (longer delays cause more advanced disease, and more advanced stages are associated with poorer survival), which means that adjusting for tumour stage will introduce spurious confounding (see Pita-Fernandez *et al.*[109]). However, Pruitt *et al.*[137] found that patients with longer diagnostic delays had earlier-stage disease, which they state is refuting the common assumption that stage is an intermediate factor in the causal chain between diagnostic delay and survival. A simpler explanation might be that the first symptom was not from the cancer, and the diagnostic interval has been artificially inflated.[139] However, the findings could also be a simple reflection of the fact that the association between diagnostic delay and survival is seriously confounded as an estimate of causal effect.

Clear definitions of diagnostic intervals

All included studies provided clear definitions of the intervals that they were evaluating. However, only one study[99] defined its interval according to the Aarhus statement.[92] Two previous studies[135,136] conducted by the same authors were published prior to the Aarhus statement.[92] Four studies[101,107,109,137] used definitions similar to those reported in the Aarhus statement,[92] and all but one study[101] referred to the time period as provider, diagnostic, or treatment 'delay', rather than using the term 'interval', which the Aarhus statement[92] recommends. Three studies[110,111,114] used definitions that reflect a national guideline target. In one study,[113] which compared the use of three different data sources, the defined TDI could have been interpreted differently according to whether the data were obtained from the hospital records, GP records or patient questionnaires.

Lead time bias or immortal time bias

The previous review[25] recommended the use of survival as the 'gold-standard' outcome measure. Methodological issues to consider when interpreting the findings include the potential effect of either lead time bias or immortal time bias. Lead time bias may be present when early detection advances what would have been the original date of diagnosis to an earlier point in time, while not necessarily delaying a patient's time of death.[99] It has been suggested that, when studying diagnostic delay, survival time should ideally be measured from the date of first symptom,[93,94] rather than the date of diagnosis, to overcome lead time bias.[99] This was done by one study;[107] however, it may not always be possible, particularly when using retrospective analyses of administrative data.[137] Furthermore, calculating survival time from the onset of symptoms includes a period before the occurrence of the defined exposure of interest (i.e. diagnosis), which may, in turn, create an immortal time bias.[99] One study[135] used data obtained from three previous studies,[81,136,138] one of which collected data on date of symptom onset using patient interview questionnaires.[138] To account for 'immortal person-time' in the patient-based study, Tørring *et al.*[135] specified delayed entry (left truncation) from the date of the interview. Pruitt *et al.*[137] used a case–control design to try to mitigate both these biases. Cases (deaths due to CRC) and controls (deaths due to other causes or censored) were matched on survival time, and the association between delay and death was examined using logistic regression.[137]

Survival as a function of age

Tørring *et al.*[99] adjusted for age at diagnosis, which they state would ensure that any increase in mortality is not a natural function of becoming older. Most of the studies included age in their multivariate analysis, but this was incorporated as a simple binary measure in some studies (≤ 69 vs. > 69 years,[108] ≥ 70 vs. < 70 years[106] and < 65 vs. ≥ 65 years[113]). Four studies adjusted for age categorised as multiple groups: 18–59, 60–74 and ≥ 75 years;[99] < 50, 50–60, 60–70, 70–80 and > 80 years;[109] ≤ 65; 66–69, 70–74, 75–79, 80–84 and ≥ 85 years;[137] and 15–44, 45–54, 55–64, 65–74 and ≥ 75 years.[101] One study[107] included age as a continuous variable.

Further potential sources of bias or confounding

Most studies were retrospective and set in linked databases. Other data sources for interval data or timing of events were GP questionnaires, patient records and interviews. Potential limitations include recall bias and validity and accuracy of the data. Another consideration is the accuracy of the data

collection conducted as part of the study. The use of multiple sources for the same data is one approach to augment or validate the data. Janssen *et al.*[114] reported that two investigators independently reviewed a subset of subject data to obtain a measure of interobserver agreement. No other study reported doing this. No other study reported using a process to account for the accuracy of interval classification. Some studies did report excluding cases with missing data or discrepant data for calculating the intervals.

The 2011 study by Tørring *et al.*[136] relied on the GPs retrospectively ascribing a date to relevant milestones on the diagnostic pathway of each patient. The authors acknowledge that these dates may have been affected by differential information bias, because of the GPs' extensive knowledge of their patients. The subsequent (2012) study by Tørring *et al.*[135] was based on three cohort studies that used different sources to ascertain the date of first presentation: GP questionnaires (Tørring *et al.*[136]), interviewer-administered patient questionnaires (Korsgaard *et al.*[138]) and primary care records (Hamilton *et al.*[81]). Leiva *et al.*[113] also compared data on intervals identified using three different sources of information (patient interviews, hospital records and GP records), but there was heterogeneity in the types of intervals studied.

Another potential issue that has implications for decision-analytic modelling is that some studies have shown that rectal and colon cancers have differing results, suggesting that these two should not be analysed together.[137] Only Tørring *et al.*[135] adjusted for the separate cancer sites.

Discussion

Summary of the findings

The current review updates the findings of a previous systematic review[25] that examined the association between different durations of time from first symptom to diagnosis or treatment and clinical outcomes across all major cancers. Summary details of included studies are presented, along with an overall assessment of whether or not each study reported an association between shorter times to diagnosis and a more favourable outcome. A more in-depth evaluation was conducted of CRC studies to inform the decision-analytic model. Summary details of included studies, including their results, are presented in structured tables in *Appendix 4*. No meta-analyses were undertaken because of heterogeneity, which included variability and how intervals were assessed. The findings of some of the more recent studies indicate that the relationship between diagnostic interval and patient outcomes is likely to be U-shaped, with both very short and long intervals associated with poor outcomes. The review also identified important biases and other factors that may affect the findings of studies in this field.

Existing reviews of colorectal cancer

Previous systematic reviews conducted by Ramos *et al.*[94,95] found no association between delays and stage at diagnosis[95] or survival[94] for CRC. The initial review by Ramos *et al.*[94] included 26 studies, 20 of which showed no associations between delay and survival. In contrast, four studies showed that the delay was a factor contributing to better prognosis, and two showed that it contributed to poor prognosis.[94] The detailed literature review by Thompson *et al.*[96] also found no strong theoretical basis for a benefit from earlier diagnosis of symptomatic bowel cancer. Five studies were identified that demonstrated improvement in survival due to early diagnosis, and 16 studies, paradoxically, showed worse outcomes associated with prompt treatment. The authors also acknowledged that it was difficult to draw firm conclusions from observational studies because of the potential biases. The overall findings of these reviews may be due to the apparent U-shaped association between the interval from first symptom to diagnosis or treatment and patient outcomes in CRC.

Strengths and limitations of the updated review

The original review,[25] which we have updated, was the first systematic review to evaluate the impact of any pre-diagnostic interval on patient outcomes in any cancer. This meant that the review included a very large number of heterogeneous studies. As a result, only a broad narrative synthesis was undertaken, with the

overall findings for all cancer sites presented in a summary table, similar to *Table 11*. The current review focused on providing an update of the previous review's overall findings, and also emphasised whether or not studies found statistically significant findings, and the direction of the effects of any such findings. The synthesis of both the previous and the current review essentially represents a 'vote counting' process, and does not take into account the magnitude of the effect (or any potential U-shaped association).

In view of the fact that only a narrative synthesis was feasible, and that the association between the interval T8 and the outcome survival was the only association to be evaluated by more than two studies, the current review of CRC studies focused on providing an assessment of the methodology of included studies; the summary results of individual studies are presented in *Appendix 4*.

One of the main limitations of relying on statistical significance to identify relevant associations is that it ignores studies that lack sufficient power to detect small differences. One of the main advantages of a meta-analysis is that it can improve power by synthesising the findings of multiple studies. Very few included CRC studies assessed the associations between the same intervals using the same outcome measure to allow any meta-analysis, and, when they did, they varied in the approach used to analyse the data; whether the interval was considered as a continuous, categorical or dichotomised variable; and the selected cut-off points. However, even if a sufficient number of homogeneous studies were available, accounting for the potential U-shaped association in a meta-analysis is not trivial. The intervals would need to be considered as a categorical or continuous variable, which may not be reported in the primary studies.

Another important limitation of the updated review is the likelihood of selective outcome reporting. Some studies reported evaluating multiple intervals but did not provide the outcome data for all of them. It is also likely that some studies did not report all the intervals they evaluated, and reported the findings for selected intervals only. Selective reporting may have been a particular issue with studies that incorporated a broader aim of evaluating more than just the association between timey diagnoses and favourable outcomes; for example, they may also have aimed to describe the diagnostic journey and identify factors associated with delay.

The literature searches for the original review identified > 193,077 references. Our updated review represents a supplementary review that was conducted to inform the decision model, and needed to be completed in a timely manner. The impracticality of screening thousands of references in a timely manner when updating any systematic reviews means that new efficient methods of undertaking this process are needed. Our updated review used an innovative and pragmatic method for searching the literature, which can contribute to the future development and availability of such methods. We tested the appropriateness of the new approach by applying it to the second phase of the previous review[25] and comparing the findings with those achieved in the original literature searches. Our novel approach did not capture all the relevant studies in this test. However, some of this may be because of the refinement in the inclusion criteria that occurred between Phases I and II of the original review. It is therefore unclear if the current review has missed any important studies using these pragmatic search methods. Further work is needed to refine and test this new searching approach.

The original systematic review[25] was undertaken in two phases: an initial review was conducted in 2010 and a subsequent review undertaken in 2013 covering the literature published since 2010. Both phases were published as a single review in 2015.[25] The original 2010 phase did not include a review of CRC studies, as there were existing systematic reviews by Ramos *et al.*[94,95] (one in 2007[94] for survival outcomes and another in 2008[95] for stage) for this cancer site. This means that the previous review[25] included primary studies for CRC published only after 2010. The 2007 Ramos *et al.*[94] review identified 12 studies that they reported as having used multivariate analysis to evaluate the association between diagnostic or therapeutic intervals and survival. However, the authors did not report which studies these were; therefore, we were unable to consider these studies for inclusion in our review as studies addressing the waiting-time paradox.

Conclusion

The overall findings of the updated review highlight the uncertainty in the evidence base for the extent to which reducing times to diagnosis and/or treatment leads to improved patient outcomes. There is still a lack of firm evidence to demonstrate an association between timely diagnosis and favourable outcomes in those with CRC. The evidence indicates a U-shaped relationship, but also reflects considerable heterogeneity in terms of the methods and intervals evaluated, and highlights some important methodological challenges in evaluating whether or not timely diagnosis leads to better outcomes. The findings support the need to expedite the diagnosis of symptomatic patients and to work to prevent the preventable delays that some patients continue to experience.

Chapter 6 Data for informing the economic decision model

Objective

The economic decision-analytic model explored the uncertainties regarding the clinical effectiveness and cost-effectiveness of the tools to aid cancer diagnosis decision-making in primary care in the NHS. Given the scope of the current study, it was not feasible to model the impact of diagnostic tools for all common cancer types for which diagnostic tools exist. Instead, at the outset, we chose to model one common cancer, CRC, to illustrate a case for which we believe there are good a priori reasons why diagnostic tools in primary care could affect patient outcomes and may be effective and cost-effective. CRC is among the cancers with the shorter diagnostic intervals (31–60 days); cancers with longer diagnostic intervals, for example bladder cancer (61–91 days) and lung cancer (91–120 days), would have greater scope for diagnosis. If clinical effectiveness and cost-effectiveness results are favourable for CRC, the modelling approach could be extended to other cancers, but this is beyond the scope of the current study.

We will use the decision-analytic model to link empirically demonstrated impacts on short- and medium-term outcomes to the effects of treatment, to estimate the impact on longer-term outcomes. In particular, we will explore the potential impacts on health service resource use, costs and patient outcomes in relation to CRC. This chapter focuses on the sources of information from previous chapters, a methodological review of the decision-economic models on CRC that have been published and the data for populating the decision-analytic model.

Sources of information

The decision-analytic model is informed by the findings from SR1 (clinical effectiveness of the diagnostic tools), the accuracy of the diagnostic tools (identified in SR2), and by the updating of a previous systematic review[25] which provides evidence on the impact of reducing time to cancer diagnosis and/or treatment on patient outcomes. Relevant studies were those that evaluated tools in UK primary care patients, and reported outcomes associated with the use of the tools in terms of referral, diagnostic or treatment intervals, stage at diagnosis, patient health-related quality of life or survival.

The model is also informed by a review of decision-analytic models that sought to identify the strengths and limitations of previous models and quality of evidence on parameter values for resource use, costs and utilities for populating a decision model of primary diagnostic strategies for suspected cancers. Because modelling the use of the diagnostic tools required a CRC disease model to translate diagnostic outcomes into clinical outcomes following treatment, and existing studies of CRC screening were known to us that included whole disease history model for the UK, we extended the scope of our review of economic studies to include the population of asymptomatic patients so that relevant disease models could be identified.

Summary of findings for the decision-analytic model

Systematic review 1

Two studies from SR1 were relevant for the CRC decision-analytic model: Hamilton *et al.*[13] and Emery *et al.*[35]

In the study by Hamilton *et al.*,[13] seven English cancer networks were selected for using the RAT (mouse mat or desktop flip chart versions) for colorectal and lung cancer, to provide the underdiagnosed cancer risk estimates in individuals aged > 40 years presenting to primary care with symptoms. The authors[13] evaluated the changes of 2WW referrals and request of colonoscopies by GPs for the two 6-month periods before and after the distribution of the RAT. A total of 614 GPs from 165 practices were recruited. The RAT for CRC was used and completed 1521 times, and was associated with increases of 26% (from 1173 to 1477) and 15% (from 1762 to 2032) in 2WW referrals and colonoscopies performed, respectively; these increases resulted in 10 new additional cases identified (from 134 to 144). Unfortunately, the authors[13] were unable to distinguish which of the additional 270 colonoscopies were ordered for patients for whom the RAT was used. However, the number of 2WW, routine and urgent referrals reportedly accounted, respectively, for 49% (702/1433), 12.7% (182/1433) and 10% (144/1433) of patients evaluated with the CRC RAT.

The study by Emery *et al.*[35] used a 2 × 2 factorial cluster RCT design in rural Western Australia. The objective of the study was to measure the effect on the TDI of a community-based symptom awareness and general practice-based educational interventions for individuals presenting with symptoms of breast, prostate, colorectal or lung cancers. For the decision model, the intervention relevant is the one at GP level, which consisted of a GP resource card with symptom risk assessment charts (RAT) and local cancer referral pathways in which the primary outcome was the duration of the TDI. As mentioned in SR1, for colorectal cancer, the authors did not find statistically significant differences in the median TDI (GP-based intervention, $n = 124$; GP control, $n = 122$) or in the log-transformed (ln) mean difference (0.3, 95% CI −0.51 to 0.45; $p = 0.42$) for the GP-based intervention versus control, nor did they find it when the factorial design was analysed by tumour group or subintervals of the TDI.

Critique of identified data

It is important to note that the data on the diagnostic accuracy for the two tools for which relevant evidence was identified, QCancer and RAT, are derived from low-risk patient populations, which NICE has previously defined as constituted by those patients presenting to primary care without symptoms requiring 2WW referral and whose symptoms place them in the positive predictive value (PPV) range of 0.1–3.0% risk band.[140] The diagnostic accuracy values for QCancer were estimated from a sample whose prevalence, as determined by a recorded diagnosis of CRC over the 2-year follow-up period after the first appearance of symptoms, was 0.16% after excluding the red-flag symptoms, in a sample of 1,223,192 patients.

The prevalence over which the diagnostic accuracy of the RAT was estimated could not be calculated from the available report.[141] However, the recommended use of the tool by its developers is for patients presenting with signs and symptoms that have a PPV of > 2% to follow the 2WW referral pathway, those with symptoms having a PPV of 1–2% to be referred for primary care investigations and those with signs and symptoms having a PPV of < 1% to be managed in primary care using the GP's clinical judgement. When patients with signs and symptoms with a PPV of ≥ 3% are excluded because they are high risk, the resulting prevalence in patients with all the symptoms considered by the RAT should fall between 0% and 3%. This would be consistent with NICE guideline[6] definitions of a low-risk population of symptomatic patients presenting to primary care, and reflect the range of risk underlying the reported sensitivity and specificity values for the RAT (Hamilton[141]).

Systematic review 2

Table 12 summarises the relevant evidence on predictive accuracy of externally validated models in UK patients, as described in *Chapter 4*. Of the four models, one reports no diagnostic accuracy data (the Netherlands model[55]), and another reports no specific diagnostic threshold to define positive findings (the BB equation[22]). A third model has no reported diagnostic accuracy estimates in UK

TABLE 12 Externally validated predictive model accuracy in UK patients for populating the decision model

Study	Number and category of descriptors	Study design	Source	Sensitivity, % (95% CI)	Specificity, % (95% CI)
BB equation	8; symptoms, test results (no threshold given)	Retrospective case–control	Marshall 2011[22] (validated in Dutch sample, Elias 2017[50])	33.5[a]	97.7[a]
Netherlands model	3; symptoms, patient demographics	Prospective cohort	Hodder 2005[77]	N/A	N/A
QCancer	6 (females) and 7 (males); symptoms, medical history, test results	Prospective cohort	Hippisley-Cox 2012[80] and Collins 2012[15]	• Women: 70.7 (70.6 to 70.8) • Men: 74.1 (74.0 to 74.2) • Top 10% risk score, excluding top 1% risk: 61%[141]	• Women: 90.3 (90.3 to 90.4) • Men: 90.2 (90.1 to 90.2) • Top 10% risk score, excluding 1% risk score: 91%

RAT	10; symptoms, test results	Case–control prospective cohort	Hamilton 2005[81] (version in Marshall 2011[26] validated in Dutch sample, Elias 2017[50])	• No report available giving a threshold score[b]
				• Unpublished doctoral dissertation:
				  ○ Sensitivity 69
				  ○ (Score of 35)
					• No report available giving a threshold score[b]
					• Unpublished doctoral dissertation:
					  ○ Specificity 77
					  ○ (Score of 35)

N/A, not applicable.
a Marshall *et al.*[22] report sensitivity and specificity values without stating the threshold.
b Marshall *et al.*[22] report sensitivity of 32.6% and specificity of 97.8% without stating the threshold used, which in any case is unlikely to have been the recommended one of 3% for the UK.

patients at a specific threshold (the RAT[74]). However, these data are available from a doctoral thesis dissertation, identified after SR2 was completed (William Hamilton, University of Exeter, 2019, personal communication), in which sensitivity of 69% and specificity of 77% are reported for a score of 35, corresponding to a threshold PPV of 2%.[141] The remaining model QCancer, provides sensitivity and specificity parameter estimates by sex, which may be used to populate a decision model of decision-making for suspected cancers. However, these values include patients with red-flag symptoms, which may be deemed appropriate for 2WW referrals, so a more appropriate value for the low-risk patient primary care population of interest to our study would exclude these cases. Therefore, we calculate the sensitivity and specificity of QCancer for the top 10% risk threshold, after excluding the top 1% risk observations (i.e. subtracting the published number of patients and CRC cases in the top 1% risk from the top 10% risk and cases total, respectively) from the published data,[77] to be 61% and 91%, respectively. These values are derived from analysis of a large database of electronic medical records of patients seen in routine practice.

Updated review

As discussed in *Chapter 5*, the evidence identified in the updated review of the relationship between diagnostic delay and cancer patient outcomes does not allow the derivation of valid estimates on the effect of expeditious CRC diagnosis on cancer stage or patient survival. In particular, the documented non-monotonic (e.g. U-shaped) relationship between diagnostic interval and mortality does not infer an effect mediated through the impact of diagnostic interval on stage of diagnosis. This is because shorter

diagnostic intervals are associated with unobserved differences in cancer stage and tumour aggressiveness, and are affected by lead time and immortal time biases whereby deaths from cancer may occur outside the observed follow-up at varying degrees across the duration of the diagnostic interval. Similarly, the observed relationship between diagnostic interval and cancer stage may be affected by unobserved variation in underlying disease severity and cancer aggressiveness across the range of the diagnostic interval.

Given the lack of valid surrogate or clinical outcome data for building an economic model of diagnostic tools or decision-aid models in primary care, our cost-effectiveness analysis was limited to an exploration of the expected effects of prolonged intervals to diagnosis associated with the relative diagnostic accuracy of diagnostic tools. This limited aim will naturally account for more sources of uncertainty than an analysis based on clinical surrogate outcomes, but reflects a realistic approach given the quality of the data available.

Systematic review of existing economic decision-analytic models for colorectal cancer

Methods

The systematic review of decision models and evidence used to populate their parameters used a rigorous but pragmatic approach because of the time constraints of developing a de novo model consistent with good modelling practice.[142,143]

Search strategy

The search strategy was designed to retrieve economic decision models in CRC. The following sources were searched: MEDLINE, MEDLINE In-Process & Other Non-Indexed Citations, EMBASE and Health Management Information Consortium (HMIC) [all via Ovid® (Wolters Kluwer, Alphen aan den Rijn, the Netherlands)], NHS Economic Evaluation Database (NHS EED) and the Health Technology Assessment (HTA) Database (both via the Cochrane Library) and EconLit [via EBSCO*host* (EBSCO Information Services, Ipswich, MA, USA)]. The searches were developed and run by an information specialist (SR) in September 2017 and updated in September 2018. A search filter was used to limit the searches to economic studies. No date or language limits were used. The database search results were exported to, and de-duplicated using, EndNote X7. De-duplication was also performed using manual checking. The search strategies for each database are detailed in *Appendix 5*. Studies were selected using predefined inclusion and exclusion criteria, which are presented in the following sections.

Inclusion criteria

The studies were included if the:

- population considered was adults (aged ≥ 18 years) with symptomatic or asymptomatic CRC, or at risk of developing CRC
- interventions included diagnosis CRC models, screening programmes to identify CRC in the population (as long as they also included a disease progression model), or treatment of CRC
- comparator was current clinical practice or other
- outcomes included resource use, costs and health outcomes [which included life-years gained, quality-adjusted life-years (QALYs) gained, CRC cases prevented, etc.]
- studies were trial- or model-based that included a CRC disease model using the Dukes' staging or the Classification of Malignant Tumours of the American Joint Committee on Cancer (AJCC), that is, a tumour node metastasis (TNM)-based classification staging system (cost-effectiveness, cost–utility) [both systems are commonly used in the UK and by NICE, although the AJCC's system is mostly known in the UK as TNM (classification of malignant tumours). Further explanation of these classification systems can be found in *Appendix 5*.].

Exclusion criteria

Studies were excluded if one or more of the eligibility criteria were not met, if the studies were cost of illness or burden of diseases, if they reported only methodological issues, or if they were abstracts, reviews, commentaries, letters or editorials.

Selection of studies

Titles and abstracts obtained from the search were screened independently by two reviewers (PL and BG) using the inclusion/exclusion criteria. All potential included studies were then screened by a third reviewer (AML). Full articles were screened for eligibility by two reviewers (PL and AML). Any disagreement was resolved by consensus.

Data extraction

Standardised data forms were used to extract the relevant information for each of the included studies. One reviewer (PL) extracted the data and the information extracted was checked by a second reviewer (AML). The data extracted included author, year, type of model (diagnostic, screening, disease progression) and study question (*Table 13*), perspective, population, outcome measure, model duration,

TABLE 13 Overview of studies included

Study	Country/region	Pathway of care (screening/ diagnostic/treatment)	Objective
Allen 2005[144]	USA	Diagnosis and natural history	To compare the cost-effectiveness of four diagnostic strategies for evaluating rectal bleeding
Tappenden 2007[145]	England	Disease progression	To develop a state-transition model to simulate the life experience of a cohort of individuals without polyps or cancer through the development of adenomatous polyps and malignant carcinoma and subsequent death in the general population
Tsoi 2008[146]	Asian countries	Screening and disease progression	To evaluate the cost-effectiveness of FOBTs, FS and colonoscopy on the basis of disease prevalence, compliance rate and cost of screening procedures in Asian countries
Zauber 2008[147]	USA	Screening and disease progression	To assess life-years gained and colonoscopy requirements for CRC screening strategies and identify a set of recommendable screening strategies
Heitman 2010[148]	Canada	Screening and disease progression	To perform an economic evaluation of CRC screening in average-risk North American individuals considering all relevant screening modalities and current CRC treatment costs
Lee 2010[149]	UK	Screening and disease progression	To assess the cost-effectiveness of three-dimensional CTC vs. OC for colonic imaging of symptomatic gastroenterology patients
Knudsen 2012[150]	USA	Screening and disease progression	To assess the clinical effectiveness and costs of colonoscopy vs. other rescreening strategies
Sharp 2012[151]	Ireland	Screening and disease progression	To evaluate the cost-effectiveness of a population-based screening programme
Whyte 2012[152]	England	Screening and disease progression	To use newly available data to estimate the cost-effectiveness and endoscopy requirements of screening options for CRC to inform screening policy in England

continued

TABLE 13 Overview of studies included (*continued*)

Study	Country/region	Pathway of care (screening/ diagnostic/treatment)	Objective
Goede 2013[153]	The Netherlands	Screening and disease progression	To evaluate if two-sample FIT screening is cost-effective compared with one-sample FIT
Gomes 2013[154]	The UK	Screening and disease progression	To assess the cost-effectiveness of CTC for CRC screening
Tappenden 2013[155]	England	Diagnosis and disease progression	To assess the feasibility and value of simulating whole disease and treatment pathways within a single model to provide a common economic basis for informing resource allocation decisions
Whyte 2014[156]	England	Awareness campaign, screening and disease progression	To estimate the clinical effectiveness of a CRC awareness campaign
Cantor 2015[157]	USA	Screening using patient decision aids	To provide a framework for analysing the cost-effectiveness of decision aid for CRC screening
Pil 2016[158]	Belgium	Screening and disease progression	To assess the cost-effectiveness and budget impact analyses of the CRC screening
Wong 2016[159]	Hong Kong	Screening and disease progression	To evaluate the cost-effectiveness, by age and sex, of CRC screening
Coldman 2017[160]	Canada	Disease progression	To develop a simulation model to predict the effect of threshold selection on outcomes including life-years gained, CRC incidence and mortality, and costs
Westwood 2017[161]	England and Wales	Diagnosis and disease progression	To assess the clinical effectiveness of FITs for primary care triage of people with low-risk symptoms

CTC, computerised tomographic colonography; FIT, faecal immunochemical test; FOBT, faecal occult blood test; FS, flexible sigmoidoscopy; OC, optical colonoscopy.

type of uncertainty analysis, discount rate, base year and prices (*Table 14*). Additional data extraction for sensitivity analyses was carried out to assess which parameters had the changes in value that most affected the incremental cost-effectiveness ratios (ICERs); this information is presented in *Appendix 5*.

Critical appraisal

The methodological quality of the studies included was assessed in detail by two reviewers (PL and AML) following the checklist for model studies by Philips *et al.*[162]

Results

Studies identified

A first electronic search retrieved a total of 2730 hits; after de-duplication, 2129 were screened, of which 2109 hits were excluded. In addition, we were aware of two key cost-effectiveness studies that were missed in the search and decided to run a new search with a more refined set of search terms (for the revised search, see *Appendix 5*) to identify more cost-effectiveness models. We assessed the full text of 20 studies from the first search, seven additional studies from the revised search and one new study from the updated search. After a more in-depth screening of these 28 studies, 13 papers were excluded: for nine papers, the disease model description was not based on the Dukes' or

TABLE 14 Economic decision-analytic models: study characteristics

Study	Perspective	Population	Diagnostic strategies	Outcome measure	Model structure	Model duration	Uncertainty analysis	Discount rate (%)	Base year and currency
Allen 2005[144]	Modified societal	Patients aged 55 years presenting one or more episodes of rectal bleeding	• Watchful waiting • FS • FS + ACBE • Colonoscopy	QALYs	Markov state-transition model, annual cycles	Lifetime	Two-way sensitivity and scenario analyses	3.5	2001; USD
Tappenden 2007[145]	Assumed to be UK NHS	General population of England aged 50–60 years	N/A	QALYs	Markov state-transition model, annual cycles	Lifetime	One-way and probabilistic sensitivity analyses	3.5	Not explicit (2004); GBP
Tsoi 2008[146]	Unclear	Aged 50 years	N/A	Life-years	Decision tree model	80 years of age	One-way and two-way sensitivity analyses	3	Not clear or explicit (1996, 2003); USD
Zauber 2008[147]	Societal	US average-risk population aged 40 years	N/A	Life-years	Microsimulation model	Lifetime	One-way sensitivity analysis	Not reported	Not reported
Heitman 2010[148]	Third-party payer	Average-risk individuals aged 50–75 years	N/A	QALYs	Markov model	Lifetime	Univariate sensitivity and probabilistic sensitivity analyses	5	2008; CAD
Lee 2010[149]	UK NHS	Aged 60–69 years	N/A	Life-years, QALYs	Markov state-transition model	Lifetime	Univariate sensitivity and probabilistic sensitivity analyses	3.5	2007; GBP
Knudsen 2012[150]	Societal	Aged ≥ 50 years, with no adenomas or cancer detected	N/A	Lifetime measures per CRC, life expectancy	Simulation model	Lifetime	One-way sensitivity analysis	3	2007; USD
Sharp 2012[151]	Third-party payer	• Aged 55–74 years • Aged 60 years	N/A	QALYs	Markov state-transition model, annual cycles	Lifetime	Probabilistic sensitivity analyses	4	2008; EUR

continued

TABLE 14 Economic decision-analytic models: study characteristics (*continued*)

Study	Perspective	Population	Diagnostic strategies	Outcome measure	Model structure	Model duration	Uncertainty analysis	Discount rate (%)	Base year and currency
Whyte 2012[152]	UK NHS	Individuals aged 30 years in the general population of England with normal colon epithelium	N/A	QALYs	State-transition model, annual cycles	Lifetime	One-way sensitivity analyses and probabilistic sensitivity analyses	3.5	Not clear or explicit (2008 or 2010); GBP
Goede 2013[153]	Health-care system	From birth to death	N/A	Life-years	Microsimulation	Lifetime	Univariate sensitivity analysis	3	2010; EUR
Gomes 2013[154]	UK NHS	Aged ≥ 50 years	N/A	Life-years, QALYs	Markov state-transition model	Lifetime	Univariate sensitivity and probabilistic sensitivity analyses	3.5	2010; GBP
Tappenden 2013[155]	UK NHS	General population, includes diagnosis in primary care	RAT	QALYs	Discrete event simulation model	Lifetime	Constrained maximisation analysis	3.5	2011; GBP
Whyte 2014[156]	UK NHS	General population	N/A	Cancer mortality, QALYs	State-transition model with annual cycles	Lifetime	Two-way sensitivity and probabilistic sensitivity analyses	3.5	2012; GBP
Cantor 2015[157]	Not explicit	General population (adults aged ≥ 50 years). Eligible for screening	N/A	Life-years	Decision tree model	Lifetime	One-way sensitivity analysis	Not reported	2013; USD
Pil 2016[158]	Societal	Adults aged between 56 and 64 years eligible for CRC screening	N/A	Predicted mortality, QALYs	Decision tree and Markov models	20 years	One-way sensitivity and probabilistic sensitivity analyses	• 3 for costs • 1.5 for health effects	2014; EUR

Study	Perspective	Population	Diagnostic strategies	Outcome measure	Model structure	Model duration	Uncertainty analysis	Discount rate (%)	Base year and currency
Wong 2016[159]	Not explicit	Asymptomatic subjects aged 50 years	N/A	Life-years	Markov model	20 years	One-way sensitivity and probabilistic sensitivity analyses	• 3 for costs • Not reported for utilities	Not clear or explicit (1996, 2003 or 2008); USD
Coldman 2017[160]	Third-party payer (NHS)	General population aged 45 years	N/A	Life-years gained, CRC incidence, CRC mortality	Simulation model	Lifetime	Sensitivity analysis (not clear)	3	2014; CAD
Westwood 2017[161]	UK NHS	Symptomatic patients aged ≥ 40 years presenting to primary care who are at low risk of CRC	1. FIT 2. FOBT 3. No triage, refer all to colonoscopy	QALYs, life-years	Markov state-transition model with annual cycles	Lifetime	Scenario analysis and probabilistic sensitivity analysis	3.5	2015; GBP

ACBE, air-contrast barium enema; CAD, Canadian dollars; EUR, euros; FIT, faecal immunochemical test; FOBT, faecal occult blood test; FS, flexible sigmoidoscopy; GBP, Great British pounds; N/A, not applicable; USD, United States dollars.

AJCC staging; three screening models did not include a full disease model; and the last excluded study was a costing study of setting up a screening programme. Reference-searching led us to three additional studies; 18 studies, in total, met the inclusion criteria for the review. Further details are presented in *Figure 4*.

Study characteristics

Table 13 provides the overview of the included models. Over 50% (10/18) of the studies were conducted in Europe (seven of these were in the UK), six in North America (four in the USA and two in Canada) and two in Asia.

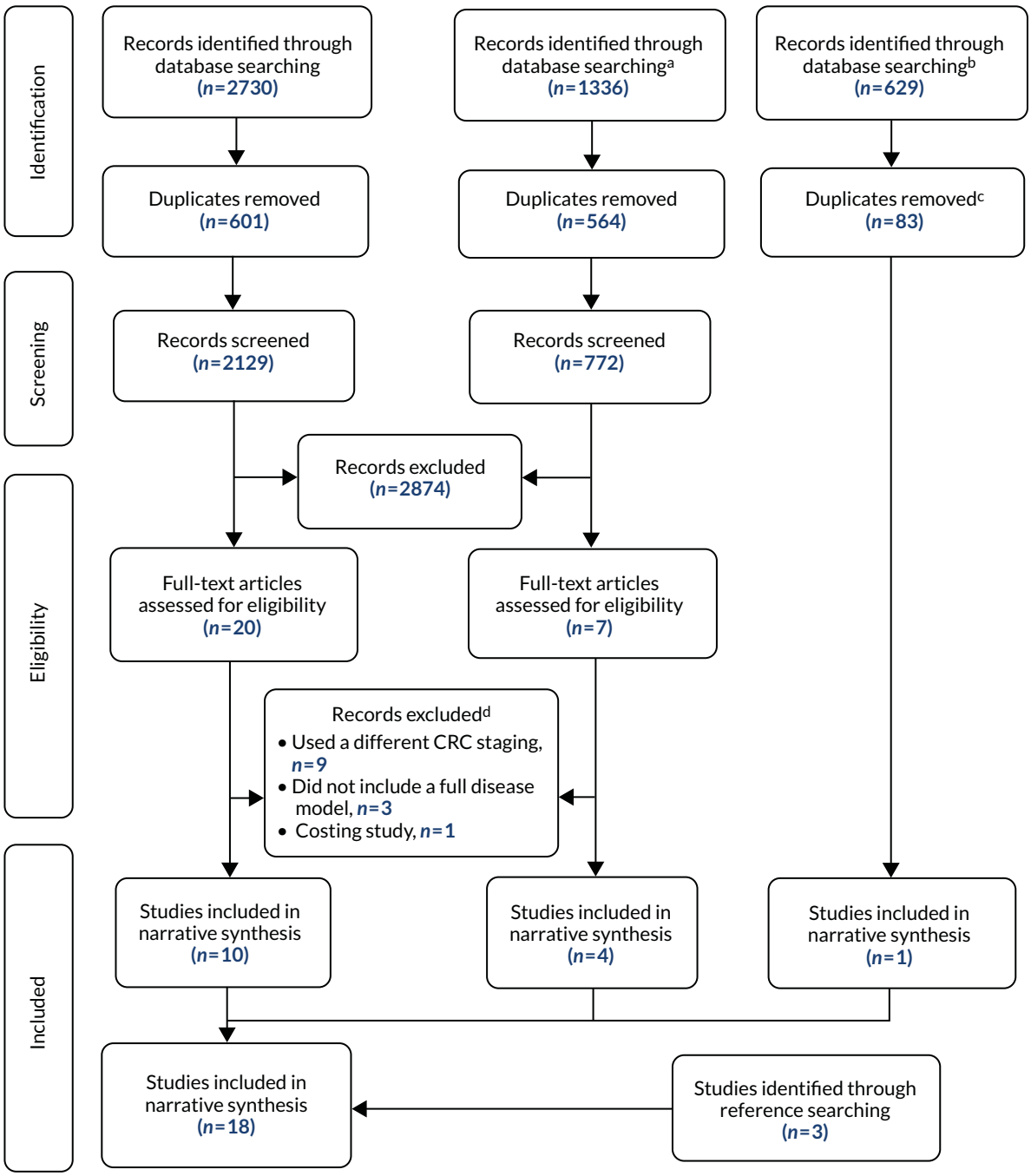


FIGURE 4 Economic decision-analytic models: the PRISMA flow diagram. a, Revised strategy; b, updated search using the revised strategy; c, a total of 546 hits were screened with the updated search; and d, further exclusion details are presented in *Appendix 5*.

Over 80% (15/18) of the studies were two-part models, mainly screening and full disease progression models (12/18); only three studies were diagnosis and natural history disease models. The other three models included in this review were disease progression models only. Two exceptions were made of studies that did not meet the inclusion criteria. The study by Cantor *et al.*[157] evaluated a patient decision aid for CRC screening; it was reviewed as it was deemed to be important for developing the diagnosis part of the model. Similarly, the model by Zauber *et al.*,[147] which excluded resource use and costs, was included despite not meeting the inclusion criteria.

In *Table 14*, the study characteristics are presented. Westwood *et al.*[161] is the only study of a diagnostic strategy in the population of interest to our study, namely a comparison of faecal immunochemical tests (FITs) at various thresholds against faecal occult blood tests (FOBTs) in a low-risk symptomatic population in a primary care setting. Of the two remaining studies of diagnostic strategies, only one[144] compared two or more diagnostic strategies, and, although it is conducted in a primary care setting, it is concerned with a high-risk symptom only (rectal bleeding in a person aged 55 years) in US primary care. The remaining study, by Tappenden *et al.*,[155] is a whole-disease model, which captures the complete pathway of CRC patients from a healthy state to clinical disease down to the detail of treatment sequences used by age and disease stage. This study's aim is not sufficiently detailed to account for the diagnostic pathway of low-risk symptoms, although it accounts for the role of a RAT in the transition from asymptomatic to clinical disease,[39] and its main interest to our purposes lies in the methods used to build the disease history model.

Further, the disease history model in the study by Tappenden *et al.*[145] is used by most of the included screening studies[145,149,154,156] summarised in *Table 14* with few modifications. In common with most models, this model used a lifetime time horizon and considered cases aged 50–60 years.

None of the identified studies evaluated a diagnostic tool. Therefore, attention will be devoted in the rest of this chapter to reviewing the suitability of identified models to accommodate any evaluation of such tools, and the quality of the available evidence synthesised by existing models to inform the assessment of tools using any suitable existing or new model.

Critical appraisal: key model features

The critical appraisal data extracted for the 18 included model-based economic evaluation studies included are presented in *Appendix 5*, and the findings are summarised in this section.

Over 60% (11/18) of the studies considered the NHS perspective and only four studies[144,147,150,158] included a societal perspective. The perspective in three studies[146,157,159] was not explicitly stated, but, given the costs included, it appears to be their respective health systems' perspectives.

There were marked differences between the models' assumptions for the long-term rate of CRC disease progression. In 40% (8/18) of the studies,[145,147,148,151,152,155,156,161] the most common assumption was to apply an annual disease progression rate, the choice of which was clearly defined by the history of the disease. The selection of key parameters was justified in all the studies, although the quality of the chosen parameters was not fully discussed in one study.[150]

Fifty per cent of the studies (9/18) used the QALY as an outcome measure; life-years were used in six studies,[146,147,150,153,157,159] and three studies[149,154,161] included both QALYs and life-years. The parameterisation into the models was described with sufficient detail and with all the relevant sources of data in most of the studies (14/18). However, the relevance of the assumptions was difficult to understand in two studies,[144,160] and how the data were incorporated into the model was unclear in two studies.[159,160]

Over 80% (15/18) of the studies addressed properly the different types of uncertainty; in particular, point estimates and values for sensitivity analyses were stated and justified in most of the studies,

except for three studies[150,159,160] in which assumptions were less clearly described. All studies except for Pil *et al.*[158] justified the values used for the sensitivity analysis.

Probabilistic sensitivity analysis was undertaken by 55% (10/18) of studies, and one study[153] justified the absence of probabilistic sensitivity analysis based on the lack of evidence on sampling variation of parameter values. In contrast, this information was vaguely stated in, or missing from, seven studies.[144,146,147,150,157,159,160] The choice of parameter distributions used was clearly reported in 50% (9/18) of the studies; however, five studies[144,154,157,159,160] omitted the information with respect to second-order uncertainty.

The internal consistency was evaluated in 10 studies only.[144,145,147–149,152,153,155,156,161] In the absence of data on key parameters, the model was calibrated against independent data in eight studies,[145,147,148,151–153,155,156] such as the rate of disease progression in asymptomatic CRC states.

Modelling approaches

Five different types of decision models were identified in the review: (1) 11 Markov state-transition models,[144–146,148,149,151,152,154,156,159,161] (2) four microsimulation models,[147,150,153,160] (3) one Markov model combined with a decision tree,[158] (4) one decision tree[157] and (5) one discrete event simulation model.[155] The colorectal health cancer states were modelled using the AJCC staging (9/18) or the Dukes' staging systems (8/18). The model diagnosis by Cantor *et al.*[157] did not consider any staging.

Markov models were preferred because of their flexibility and simplicity in populating and calibration processes. Specifically, the transition within health states can be easily represented by a cohort in a Markov model. Given the limited number of data on key disease history model parameters, a Markov model may be more suitable to synthesise the clinical and economic evidence for evaluating diagnostic, screening and surveillance strategies in CRC than other types of models such as microsimulation and discrete event simulation models. The only study that used a discrete event simulation model used such an approach to accommodate the distribution of time to event survival and disease progression outcomes of CRC treatment, which were found to have limited importance in the results of the only diagnostic study of a low risk of suspected cancer in primary care (Westwood *et al.*;[161] see *Appendix 5*). Likewise, the three studies[150,153,160] that used microsimulation models used such an approach for its convenience to account for the interactions between individual characteristics and risk factors of the screening tests and combining cohorts with different characteristics to produce results averaged across subgroups of the patient population.

Detailed review of individual models

UK models

Tappenden *et al.*[145] constructed a model for England incorporating three elements: (1) a screening intervention model that included subsequent colonoscopy surveillance; (2) a state-transition model that simulated the natural history of colorectal neoplasia from normal epithelium, to adenoma and carcinoma sequence; and (3) a mortality model that considered age-specific other causes mortality, CRC mortality and mortality from perforation due to endoscopic procedures. Transitions between model health states were calculated using an annual cycle length until the entire model cohort was absorbed into the dead state. The model assumed that the incidence of adenomas was 1.60% and that the incidence of cancer from adenomas was 3.26%. Five screening options versus no screening were evaluated: biennial FOBTs, for people aged 50–69 and 60–69 years; two once-only flexible sigmoidoscopies, for people aged 55 and 60 years; and the fifth option was once-only flexible sigmoidoscopy for individuals aged 60 years, followed by biennial FOBTs for individuals aged 61–70 years.

The UK model developed by Lee *et al.*[149] was a two-part model: (1) the evaluation of screening strategies for CRC and (2) a state-transition Markov model representing the natural history of colorectal neoplasia (from normal colorectal epithelium to adenoma–carcinoma sequence, to death).

The four screening strategies evaluated were (1) FOBT every 2 years, (2) flexible sigmoidoscopy every 10 years, (3) optical colonoscopy every 10 years and (4) computerised tomography colonography (CTC) every 10 years. The authors used data from the study by Tappenden *et al.*[163] and recalibrated the data to adjust for estimates from recent observational data on CRC incidence, mortality and staging of cancer at time of diagnosis. The progression from normal epithelium to low-risk adenoma, low-risk adenoma to high-risk adenoma, and high-risk adenoma to Dukes' stage A CRC was 1.2%, 2.4% and 3.4%, respectively. The mortality for other causes was modelled as an age-dependent probability based on UK life tables,[164] and the probability of dying as a result of endoscopic perforation from either the screening procedure or the polyp removal (polypectomy) was included in the model as a potential complication of the interventions.

The two interlinked models developed by Whyte *et al.*[152] had (1) a model that described the screening interventions and surveillance of CRC (from invitation to screening, screening test, follow-up, diagnostic tests and surveillance) and (2) a natural history of colorectal neoplasia model (from normal colorectal epithelium to adenoma–carcinoma sequence to death). The screening intervention used the following diagnostic strategies: (1) no screening; 2) biennial guacal faecal occult blood test (gFOBT) at 60–69 years; (3) biennial gFOBT at 60–74 years; (4) immunochemical faecal occult blood test (iFOBT) at 60, 65 and 70 years; (5) biennial iFOBT at 60–69 years; (6) biennial iFOBT at 60–74 years; (7) flexible sigmoidoscopy at age 55 years; (8) flexible sigmoidoscopy at age 55 and 65 years; (9) flexible sigmoidoscopy at age 55 years and biennial gFOBT at ages 66–74 years; (10) flexible sigmoidoscopy at age 55 years and biennial iFOBT at ages 66–74 years; (11) flexible sigmoidoscopy at age 55 and iFOBT at ages 60, 65 and 70 years; (12) flexible sigmoidoscopy at age 55 years and biennial iFOBT at ages 60–74 years; and (13) flexible sigmoidoscopy at age 55 years and biennial iFOBT at ages 56–74 years. The authors assumed that probabilities associated with the transition to low-risk adenomas, the transition from low- to high-risk adenomas and the transition from high-risk adenomas to Dukes' stage A CRC were age dependent.

An interconnected model developed by Gomes *et al.*[154] comprised (1) diagnostic test and the subsequent adenoma surveillance model for symptomatic patients in secondary care, and (2) a Markov model, based on CRC natural history in the UK simulating the development of CRC in the population (from normal colorectal epithelium to adenoma–carcinoma sequence to death). The model assumes a variable annual rate of polyp development of 1.9% to 3.3% and an annual incidence of CRC from adenomas of 3.4%. The model evaluated the impact of three-dimensional CTC versus optical colonoscopy to improve the detection of adenomas and CRC.

Tappenden *et al.*[155] developed a patient-level model that simulates the disease progression and treatment pathways from pre-clinical disease through detection, diagnosis, adjuvant/neoadjuvant treatments, follow-up, curative/palliative treatments for metastases, supportive care and eventual death. The authors assumed a detailed age-specific incidence of CRC and adenoma prevalence based on Bayesian Markov chain Monte Carlo methods. The whole-disease model considered all the processes that are linked to CRC from screening, surveillance and treatment.

The model developed by Whyte *et al.*[156] comprised (1) the evaluation of the effect of a campaign to increase the awareness of the signs and symptoms of CRC and encourage self-presentation to a GP, and (2) a model to represent the development of CRC in the population (from normal colorectal epithelium to adenoma–carcinoma sequence to death). The authors assumed an increased presentation rate of 10% in each CRC Dukes' stage and an age-dependent incidence of adenomas (low and high risk) and CRC.

Westwood *et al.*[161] constructed a model for England incorporating three elements: (1) a decision model reflecting the diagnosis of CRC, (2) a Markov state-transition model to estimate the long-term costs and effects related to treatment and progression of CRC and (3) a Markov state-transition model to

estimate effects for patients without CRC. The transitions between the model health states of the two Markov models were calculated using an annual cycle length until the entire model cohort was absorbed into the dead state. The model assumed the incidence of adenomas and the incidence of cancer (from adenomas reported in Tappenden *et al.*[145]) to be 1.60% and 3.26%, respectively. The model evaluated three options of symptomatic individuals that presents to primary care: (1) FITs, (2) gFOBTs or (3) no triage tests at all (referral straight to colonoscopy).

European models

The Irish study by Sharp *et al.*[151] was a Markov state-transition model with three interlinked components: (1) the impact of screening and subsequent adenoma surveillance, (2) a natural history of colorectal neoplasia model (from normal colorectal epithelium to adenoma–carcinoma sequence to death) and (3) the impact on mortality. The authors assumed that 14% of individuals progressed from normal epithelium to stage I cancer. The second model component used gFOBT, FIT and flexible sigmoidoscopy for screening, and the diagnosis used the sensitivity and specificity of colonography and CTC. Finally, mortality in the model was assumed to be from CRC, endoscopic bowel perforation or other causes; other causes were obtained from Irish life tables and were modelled to be age dependent within each Markov cycle. The risk of dying from endoscopic perforation was considered only when using flexible sigmoidoscopy, diagnostic investigations and adenoma surveillance. The model assumed that individuals had a higher probability of dying if they had a more advanced cancer stage.

The work by Goede *et al.*[153] was based on the MIcrosimulation SCreening ANalysis (MISCAN) microsimulation model, which had three integrated components: (1) a model evaluating the screening strategies of one-sample and two-sample FIT screening with variable intervals, age ranges and cut-off levels (age at which to start screening: 45, 50, 55 and 60 years; age at which to stop screening: 70, 75 and 80 years; and screening interval at 1, 1.5, 2 and 3 years); (2) the natural history component part that simulates the development of CRC in the population (from normal colorectal epithelium to adenoma–carcinoma sequence to death); and (3) a demography component that simulates individual life histories without CRC to form a population. Authors assumed an age-dependent adenoma incidence and adenoma progression, whereas the CRC incidence rate was based on the observed incidence rate in the Netherlands.

The model developed by Pil *et al.*[158] comprised two interconnected submodels: (1) a decision tree to simulate the screening process based on FITs and (2) a state-transitional Markov model simulating the natural progression of the disease (from normal colorectal epithelium to adenoma–carcinoma sequence, diagnosis and treatment, follow-up to death). The authors assumed a sex- and age-dependent adenoma prevalence varying from 1.6% to 6.01%, and a CRC prevalence ranging from 0.11% to 0.84%.

North American models

Allen *et al.*[144] used a Markov decision model composed of three elements: (1) a model to evaluate four diagnostic strategies of a patient with rectal bleeding, (2) a model to simulate the natural history of colorectal neoplasia for patients with rectal bleeding (from no serious pathology to adenoma–carcinoma sequence and surveillance to death) and (3) a mortality model to estimate mortality among patients with CRC. The authors assumed that the progression of smaller adenomatous polyps to large polyps would take an average of 10 years and that the progression of large adenomatous polyps to invasive cancer would also take an average of 10 years. Once invasive cancer was detected, the average transition of 2 years, 1 year, and 1 year for Dukes' stage A to B, B to C, and C to D staging, respectively, was assumed. It was also assumed that 7% of individuals progressed from adenomas to Dukes' stage A and that the prevalence of adenomas was 16%. The diagnostic model evaluated (1) watchful waiting, (2) flexible sigmoidoscopy, (3) flexible sigmoidoscopy with consequent air-contrast barium enema and (4) colonoscopy. Finally, the mortality estimates used within each stage of cancer were obtained from the Surveillance, Epidemiology, and End Results (SEER) cancer registry.

Zauber *et al.*[147] used two previous developed microsimulation models [Simulation Model of Colorectal Cancer (SimCRC) and MISCAN] to reproduce the natural history of colorectal neoplasia (from no lesion to adenoma–carcinoma sequence to death). The semi-Markov microsimulation model simulated the effect of screening and other interventions on the incidence and mortality of CRC. It is a hybrid model, made of a Markov model and a discrete event simulation that described the progression of the underlying colorectal disease (e.g. from adenoma to carcinoma sequence) in an unscreened population. Authors assumed an age-dependent variability of adenoma prevalence varying from 10.2% to 36.7% and a CRC incidence varying from 5.3% to 7.3%. The model evaluated the following screening strategies: (1) no screening, (2) colonoscopy, (3) FOBT, (4) flexible sigmoidoscopy with biopsy and (5) flexible sigmoidoscopy combined with FOBT. For each strategy, individuals were evaluated at the following ages: 40, 50, 60, 75 and 85 years. FOBT strategies considered screening intervals of 1, 2 and 3 years; for the sigmoidoscopy and colonoscopy strategies, the intervals considered were 5, 10 and 20 years. The prevalence of pre-clinical cancer and adenomas was assumed to be zero.

The two-part model in Heitman *et al.*[148] included (1) the impact evaluation of screening strategies on average risk in a North American population and (2) a Markov model reproducing the natural history of the colorectal neoplasia (from normal colon to non-advanced/advanced adenomas, to CRC, to death). The authors assumed a prevalence of 17% of non-advanced adenomas, 3.8% for advanced adenomas and 0.1% for CRC. The screening strategies evaluated were (1) gFOBT or FIT annually, (2) faecal deoxyribonucleic acid (DNA) test every 3 years, (3) flexible sigmoidoscopy or CTC every 5 years and (4) colonoscopy every 10 years. The mortality rate was derived from an age-dependent population from Canada and the CRC mortality rates observed for patients with CRC according to their stage at diagnosis.

Knudsen *et al.*[150] developed a microsimulation model based on a previous study[147] (SimCRC) to represent the natural history of colorectal neoplasia (from normal colon to non-advanced/advanced adenomas, to CRC, to death) and the relative effects of screening strategies. The prevalence, size, location and multiplicity of adenomas and the incidence of CRC was assumed from the SEER programme. The model evaluated five rescreening strategies for individuals with a negative colonoscopy result at 50 years of age: (1) no further screening, (2) continuing colonoscopy every 10 years, (3) rescreening with annual highly sensitive guaiac-based FOBT, 4) annual FIT or (5) CTC every 5 years. Rescreening was assumed to begin at 60 years (10 years after the negative colonoscopy result) for all strategies.

Cantor *et al.*[157] developed a decision-analytic model comprising a decision tree to evaluate the impact of a decision aid for CRC screening. The study does not present a disease model and there is no indication of patient characteristics. The model evaluated two strategies: use of decision aid and no use of the decision aid. The decision aid supported the selection of the patients who would undergo screening. Each patient undergoing screening can have one of three diagnostic tests: (1) flexible sigmoidoscopy, (2) FOBT or (3) colonoscopy, or not undertake screening. The study assumed a change in the preferences of the screening, with a consequent increase in the number of tests, number of patients undergoing the screening and costs.

The model by Coldman *et al.*[160] (OncoSim-CRC) comprised (1) a model to predict the outcomes of biennial screening in a cohort for eight FIT threshold values of between 50 and 225 ng/ml and (2) a model simulating the natural history for the development of CRC (from normal colorectal epithelium to adenoma–carcinoma sequence, to death). The authors assumed that the prevalence of adenomas was age, sex and size dependent. The age- and sex-specific CRC incidence and mortality rates were derived from the Canadian population.

Asian models

Tsoi *et al.*[146] developed a Markov model comprising three submodels that evaluated which primary screening procedure should be adopted to reduce CRC incidence. The diagnostic tests considered were (1) FOBT, (2) flexible sigmoidoscopy and (3) colonoscopy. The annual mortality rates of CRC diagnosed

at different stages were based on the Hong Kong Cancer Registry,[165] and the mortality due to perforation of endoscopic tests was considered to be 10%. The authors assumed a prevalence of 10% for the first stage of CRC and, that 14% of individuals progressed from normal epithelium to stage I cancer.

Wong *et al.*[159] used an approach similar to that of Tsoi *et al.*,[146] but focused on developing a Markov process model to evaluate five screening strategies: (1) flexible sigmoidoscopy as a primary screening test every 5 years at the ages of 50, 55, 60, 65 and 70 years; (2) colonoscopy as a primary screening test every 10 years at the ages of 50, 60 and 70 years; (3) flexible sigmoidoscopy for each female subject every 5 years at the ages of 50 and 55 years; (4) flexible sigmoidoscopy for each female subject every 5 years at the ages of 50, 55, 60 and 65 years; and (5) flexible sigmoidoscopy for each female subject every 5 years at the ages of 50, 55, 60, 65 and 70 years. The age-stratified incidence of CRC of the Hong Kong population was based on reports from the Hong Kong Cancer Registry.

Model population and time horizon

UK models

In the models developed by Tappenden *et al.*,[145] Lee *et al.*[149] and Whyte *et al.*,[156] individuals were aged 30 years. In the model developed by Westwood *et al.*,[161] individuals entered the model at age 40 years, and in the models developed by Whyte *et al.*[152] and Gomes *et al.*,[154] individuals were aged 50 years. The only study that included individuals from birth is Tappenden *et al.*[155] All of the UK models followed individuals until death.

European models

In line with the UK models, in two of the European models, by Sharp *et al.*[151] and Pil *et al.*,[158] individuals entered the model at the ages of 30 and 50 years, respectively. However, only the model developed by Sharp *et al.*[151] followed up individuals until death, whereas, in the model developed by Pil *et al.*,[158] individuals were followed up until the age of 70 years. In comparison, the study by Goede *et al.*[153] modelled the age distribution of the Dutch population in 2005 and all the cohort individuals were followed up until death.

North American models

In two of the US models,[144,147] the cohorts entered at age 40 years and were followed up to the age of 100 years or death. However, Heitman *et al.*[148] and Knudsen *et al.*[150] considered older cohorts (aged 50 years); both studies followed up individuals to death. Although the cohort of the Canadian model[148] entered at the age of 45 years and was followed until death, Cantor *et al.*[157] did not specify any age for their cohort.

Asian models

In both Asian models, the cohort entered at the age of 50 years, but the model developed by Wong *et al.*[159] followed up individuals to 70 years of age and the model by Tsoi *et al.*[146] followed up individuals to 80 years of age.

Critique

For the purposes of the current research questions, the review has helped to identify the Westwood *et al.*[161] model as a useful precursor for the diagnostic phase. This model is important from two perspectives: first, it is the only evaluation of primary diagnostic tests for CRC identified by our review, and, second, the population included are the low-risk symptomatic population of interest. The focus of the critique will therefore be its methodology, as well as the source of the evidence used to populate its key parameters to inform the approach taken in the novo model. In particular, we will review the mechanisms used to link the diagnostic pathway to the Markov model of disease treatment used by Westwood *et al.*[161] in terms of their suitability for incorporating a diagnostic phase that has enough granularity to allow an evaluation of costs and health benefits of tool-based diagnostic strategies.

The analysis by Westwood *et al.*[161] informed the NICE Diagnostic Guideline Number 30[140] and assessed the relevant low-risk population for using the diagnostic tools of interest to our study. A strength of the model is that it considers an intervention, FIT, that has now become the recommended standard practice in the UK. The model has policy appeal as it considers the option of referring all symptomatic patients directly to secondary care, a natural option for a country where policy-makers are concerned about expediting access to diagnosis and treatment as a way to improve cancer survival outcomes.

An attractive feature of the Westwood *et al.*[161] model is that it shows that detecting more CRC cases with diagnostic tests comes at the cost of imposing more unnecessary use of health-care resources and health risks from invasive investigations to the great majority of patients who do not have CRC. When the costs and life-years are calculated over the denominator of all patients presenting with symptoms of suspected CRC, the importance of the trade-off between sensitivity and specificity becomes apparent. The costs of colonoscopies and their associated risk are modelled in the Westwood *et al.*[161] model to explore the trade-offs between gains in sensitivity of diagnostics in primary care at the expense of exposing more false positives to invasive tests. Our de novo model analysis will follow a similar approach, which will serve to highlight that gains in sensitivity of diagnostic strategies are achieved at the expense of their specificity, especially among younger patients who are at lower risk of CRC. When modelling the impact on colonoscopies, Westwood *et al.*[161] make the simplifying assumption that the costs of CRC treatment are not linked to the clinical outcomes, but this simplification does not affect the results in a significant way because of the low proportion of cancer patients.

A limitation of the Westwood *et al.*[161] model is the quality of the evidence on the diagnostic accuracy of FITs, which was obtained from high-risk patient populations included in studies that, for the most part, were not conducted in a primary care setting. This weakness is problematic because the values of FIT of 10 µg haemoglobin (Hb)/g faeces, the recommended new standard, has sensitivity values close to 100%, which effectively limits the ability of any diagnostic tool to add to the primary care diagnostic yield (see *Chapter 7*). Furthermore, although referring all patients to secondary care may be the least risky option, the low prevalence of CRC means such a strategy is unaffordable and is not used in practice.[140]

In terms of methods, the analysis by Westwood *et al.*[161] adopts a linked-data approach, whereby survival outcomes of FOBTs are driven by the ability of the test to correctly identify patients with CRC before the disease has progressed to advanced disease stages and become less responsive to treatment. The model uses disease progression data from the disease history model developed by Tappenden *et al.*[145] and Whyte *et al.*[156] to determine the extent to which delays in referral from lack of diagnostic accuracy would affect long-term life expectancy. The limitation of this approach is the high degree of uncertainty in the rates of disease progression underlying any such analysis, which were derived from calibrations of the disease history model to observed data on CRC incidence and distribution of stage at diagnosis. We adopt a similar approach in our model, given the lack of more reliable data on the relationship of delayed diagnosis and stage at diagnosis, but raise this weakness as an area that needs further research.

A further limitation of the diagnostic phase of the Westwood *et al.*[161] model is the way the time to referral is analysed. In particular, the delays incurred by false-negative cases are not explicitly modelled in terms of diagnostic accuracy of the tests being compared, but simply assumed to last for an arbitrary amount of time, which, in their base-case analysis, was 6 months. This adds a high degree of uncertainty in the absence of any data used by the authors to justify the extent of the delay. We sought to address this limitation by building a de novo model of the diagnostic pathway that was fully consistent between the diagnostic accuracy and the number of visits and the referral interval.

In terms of the granularity of the model used to link the diagnostic pathway to the Markov model of disease treatment, the Westwood *et al.*[161] model uses an annual cycle length to model health transitions. A limitation of an annual cycle is that it may not have sufficient granularity to measure the predicted outcomes, especially given small differences involved in accuracy and referral intervals

between strategies. We addressed this limitation in the de novo model by using a 28-day cycle to link the diagnostic pathway to the Markov disease model.

The Westwood *et al.*[161] model did not account for the short-term survival benefits of reducing the time to referral and treatment, as symptomatic patients with CRC who are undiagnosed, and thus untreated, are subject to increased death risks, relative to those who have been diagnosed and started treatment, especially in disease stages II/III (Dukes' stage B/C). We addressed this limitation by building a de novo model of the diagnostic pathway that accounts for the short-term survival benefits of reducing the time to referral and treatment.

The effect of delays in diagnosis on the mental health of a patient with CRC is also omitted from the Westwood *et al.*[161] model. No evidence on the magnitude of such an impact was found in the literature reviewed for this study, and the studies reviewed failed to explore the importance of that effect for the cost-effectiveness of diagnostic strategies in primary care. Our de novo model, therefore, is unable to incorporate these effects. Measures of health-related quality of life among undiagnosed and untreated CRC patients are lacking, and are difficult to measure because of the very nature of studying undiagnosed patients.

Conclusions

The review highlights differences between existing economic models of cancer diagnosis and screening. A common approach to the problem of lack of evidence on clinical outcomes of diagnostic and screening strategies has been to model long-term outcomes of CRC using a model of CRC disease progression in asymptomatic disease obtained by calibrating a disease history model to data on CRC incidence and disease stage at diagnosis.[145,147,148,151,152,155,156,161] We applied the same approach in our de novo model, given that the updated review provided no direct evidence on clinical outcomes of diagnosis, but we recommend that further research on this issue is undertaken.

Our review found no evidence on the cost-effectiveness of diagnostic tools for managing patients in primary care with suspected CRC, but identified one study of FITs[161] in the low-risk population of interest that modelled the diagnostic phase. The model explored the trade-offs between gains in sensitivity of diagnostics in primary care at the expense of exposing more false positives to invasive tests, and is a useful precursor for the diagnostic phase of our de novo model. In addition, our critique of the Westwood *et al.*[161] model identified three areas in which we could make improvements to the diagnostic phase:

1. Time to referral – time to referral will be explicitly modelled to be consistent with diagnostic accuracy of the respective test and the distribution of the number of visits before referrals in electronic health records or retrospective reports from surveys of patients' experiences. This will include modelling delays incurred by false negatives. By coherently modelling the sensitivity of primary care diagnostic strategies and the delay in referrals and treatment, the uncertainty in relative clinical effectiveness between diagnostic strategies is reduced relative to arbitrary estimates of diagnostic delay used by the existing model in this area.
2. Cycle length to model health transitions – our model used a 28-day, rather than annual, cycle to get accurate measure of predicted outcomes, especially because of the small differences involved in accuracy and referral intervals between strategies.
3. Cancer-related mortality – the short-term survival benefits of reducing the time to referral and treatment will be explicitly modelled.

The review of existing models also highlighted that the health-related quality-of-life outcomes in the diagnostic phase have received very limited attention. Our de novo model is similarly unable to explore the health-related quality-of-life outcomes in the diagnostic phase, as much of the policy focus has been devoted to improving cancer survival.

Chapter 7 Economic decision-analytic model for colorectal cancer

Aim and objectives

The earlier chapters highlight the uncertainty in the evidence base for decision tools, which make it difficult to assess the cost-effectiveness of the use of risk tools. In this chapter, we use a decision-analytic model to demonstrate the uncertainty inherent in the current evidence base, and show the probable impact that use of the tools in clinical practice may have on patient outcomes and NHS resources. The primary role of the decision modelling is not the estimation of the single most likely point estimates of costs per QALY associated with each diagnostic tool. Instead, the decision-analytic model and the evidence that is available are used in this chapter to explore the probable range of costs per QALY, and to ask questions about the likely impact of the diagnostic tools, given the current evidence base.

The objectives of the decision-analytic model were to examine the following questions:

- What are the possible impacts on patient quality of life or survival if the diagnostic tools reduce time to diagnosis?
- Will the benefit to cancer patients who are identified earlier by diagnostic tools outweigh any disutility in extra patients who do not have cancer being referred for further investigation?
- How big an improvement in quality of life would be needed to warrant the use of these tools if there are no survival impacts associated with the diagnostic tools? Would this quality-of-life improvement be justifiable given the evidence we have?
- Could a cancer diagnostic tool be considered cost-effective if it reduces the period of extreme anxiety for patients by expediting investigation and management?
- Where are the gaps in the evidence base and where is more research needed?

Furthermore, we developed a model that anticipates evidence development and can be used alongside any studies measuring impact on patient outcome directly in the future to explore implications for cost-effectiveness. We will also be able to use the model to identify the parameters that contribute most to the overall decision uncertainty about the cost-effectiveness of decision tools and where additional research might be targeted using expected value of partial perfect information.[166–168] If considered effective and cost-effective, now or in the future, the model could also be developed to assess the budget impact of introducing cancer diagnostic tools in different populations.

It was not feasible to model the impact of diagnostic tools for all common cancer types for which diagnostic tools exist. Instead, we chose to model one common cancer: CRC.

Colorectal cancer was chosen, based on the following criteria, to provide a best-case example of how the diagnostic tools could affect patient outcomes:

- Tools exist – to compare diagnostic tools with each other and with no tool, we need diagnostic tools to be available for the specific cancer type. There are existing tools for CRC, as identified in SR2. In particular, there are RATs[75,81] and QCancer[80] for CRC, which are now within GP systems and, therefore, potentially available for GPs to use.
- Common cancer – CRC is the third most common cancer in the UK, accounting for 11% of cancers in women and 13% in men.[169]

- Whole-disease models exist – it is important that whole-disease models exist, and, although not made available by the original authors, could be replicated by published methods and adapted to incorporate the diagnostic pathways of interest.
- Wide agreement on patient management –it will be important that there is wide agreement on the treatment and management of individuals to minimise uncertainties elsewhere in the clinical pathway. CRC is a cancer for which there is wide agreement on the treatment and management of individuals.
- Evidence that use of diagnostic tools changes practice – to explore the impact of the tools, it will be important to identify a cancer type for which there is existing evidence to show where the diagnostic tools will impact, and what that impact will be. Only a few studies assess the impact of RATs in SR1 that were related to CRC. These studies reported that increased cancer referrals and investigations for CRC were associated with use of tools.[13,35]
- Evidence that change in practice affects patient outcomes – it is not enough that increased referrals and investigations are associated with use of the tools; there also needs to be existing evidence of the impact of earlier diagnosis on patient outcomes. This might include earlier stage at diagnosis, higher resection rates or improved survival, as well as improved quality of life of patients. There is some evidence to show that long (and very short) diagnostic intervals are associated with increased mortality in CRC.[99,135,137] However, as indicated in *Chapter 5* and summarised in *Chapter 6*, *Updated review*, there is great uncertainty associated with estimates of the effect of expeditious CRC diagnosis on cancer stage or patient survival.

Although the implications of prolonged time to tests, diagnosis and treatment vary by cancer type, among the low-risk symptoms there are many similarities in the diagnostic options within primary care (i.e. refer all, watch and wait, refer for further tests). Therefore, the model developed can be viewed as a template for modelling the effect of diagnostic tools in cancers other than CRC, and can be revised to include prevalence and test sensitivity/specificity and costs specific to other cancers.

Methods

Conceptualisation of the model

Patient and public involvement

Involving members of the public in health research projects is becoming standard practice, but there are fewer reported examples of patient and public involvement (PPI) in health economic or modelling studies.[170,171]

At the outset of putting together the diagnostic pathway of the model, we organised a workshop to meet with patients who have experienced bowel cancer diagnosis to seek their advice on our understanding of the current diagnosis pathway and to listen to their experiences of being diagnosed. This was facilitated by Dr Emma Cockcroft from the PPI team from the Peninsula Collaboration for Leadership in Applied Health Research and Care, which is based at the University of Exeter Medical School. This part of the study did not require separate ethics approval because it was underwritten by the current ethics approval granted to the PPI group at the University of Exeter.

Dr Cockcroft wrote an invitation to the Bowel Cancer West cancer treatment centre in Plymouth to participate in the PPI, which was published on their Facebook page (Facebook, Inc., Menlo Park, CA, USA). After several weeks, two patients contacted us and we opted to have a session in which we listened to their experiences and journeys through the diagnosis pathway and asked them to tell us their opinion about our diagrammatic representation and understanding of the diagnostic pathway. We then asked their opinions about the diagnostic pathway within primary care and our diagrammatic representation of this.

Two individuals diagnosed with bowel cancer, who were brother and sister, attended the meeting. Their bowel cancers were diagnosed through a test to assess whether or not they had inherited the gene after their father passed away from bowel cancer. Both diagnoses occurred at secondary care, so neither of them had any experience of diagnosis practice in primary care. Although at this point we knew that we would not obtain information on the diagnostic pathway in primary care, we decided to take this opportunity to discuss our understanding and modelled representation of the diagnostic pathway. After listening to their descriptions of the pathway and their very difficult experiences with multiple operations, we asked them if they could identify any potential misrepresentation of reality in an animated presentation covering the progression through the diagnostic pathway in stages, from the patient attending general practice owing to experienced symptoms to the moment when the GP decides to refer the patient to secondary care for a colonoscopy. We received very encouraging, although limited, feedback on our diagnosis pathway.

Other letters of invitation to Bowel Cancer West were published via Facebook, but, after weeks of not receiving any messages, we decided to focus on the other key stakeholder, the GP.

We opted to conduct this process in two steps. First, we opted to consult with a recently retired GP from Somerset to assess whether or not our explanations of key economic concepts and our diagnostic pathway were clear and suitable for presentation to a wider number of health professionals. As a warming-up task, a short interview was carried out in which we focused on the number of years of practice by the respondent and her experience with cancer diagnosis, that is number of patients diagnosed per year, some issues with cases diagnosed at late stage and similar topics, which allowed us to build a clear picture on the process and interactions taking place at the general practice involving symptomatic patients with suspected cancer.

What we learnt from this interview corroborated the evidence available:

- too many guidelines, so it is difficult to keep all of them in mind
- very few patients per year picked up at early staging in primary care
- symptoms easily wrongly attributed to other diseases, mainly in those aged > 65 years with comorbidities.

We then moved to have a more in-depth conversation of the diagnostic pathway with the same retired GP. However, we changed our prior strategy and let her tell us the different steps that a GP would take if they suspect that a patient may have bowel cancer. When the reasoning behind the choices was unclear, we asked for further clarification. Notes were taken that were compared with our preconceived diagnostic pathway, thereby permitting us to have a better understanding of the current pathway for bowel cancer at general practices.

A third workshop was organised at a local general practice in Exeter. We prepared a modified diagnosis pathway, constructed using the current NICE guidelines[172] and complemented with the strategies learnt from the previous interactions with the patients and the retired GP. To this diagnosis pathway we added the strategy of the 'diagnostic tool'. The diagnostic pathway was printed on colour A7 pages that we brought with us for discussion with the GPs.

All the practice's GPs ($n = 7$) attended the workshop, as well as two nurses, and the group discussion was transcribed. The GPs had very different levels of experience in terms of number of years of practice. Some of their opinions were as follows:

> *Tools would be useful if you could use as a justification for referral – but currently still have to meet NICE criteria to get onto 2WW. The criteria are 'rigid' and some referrals might get bounced back if the results from a blood test are borderline.*

We use a 'tool' which is a calculated risk in our head.

Consider litigation – normally more likely to get sued for not doing something.

Borderline' cases often bounced back by secondary care – they would need to accept the 'clout' of the diagnostic tool/aid.

We also asked them if there was any reason why they would not consider using the tool. The answers were as follows:

Stuck in ways.

Doing similar anyway – just in heads.

Don't have the tools 'to hand'.

We asked in what circumstances they would consider using the diagnostic tool:

To verify the 'gut feeling' was true, not only for referral, but for reassurance, when deciding against urgent referral or more tests, etc.

A follow-up question was if the diagnostic tool would help them to decide whether or not a patient needed referral:

Seemed this would only be the case if a tool had NICE backing.

After these questions, all the different strategies in our diagnosis pathway were discussed, which resulted in *Figure 5*. An additional one-to-one meeting was held with one of the GPs to ask for further clarifications.

The diagnostic pathway

A decision-analytic model was used to investigate the key areas of uncertainty in the economic evaluation of primary care tools for informing decisions on referrals of suspected cancer patients from primary to secondary care in England.

The analysis compares costs and clinical outcomes between a diagnostic strategy whereby primary care doctors use the diagnostic algorithm in conjunction with information from investigations in primary care to decide whether or not a patient presenting with low-risk symptoms needs to be referred to secondary care for further investigations, and a strategy of primary care investigations alone. We compare the intervention strategy of RATs (CAPER)[75] or QCancer[80] as a triage to primary care investigations (which, in the model, is represented by FITs) or direct referral to secondary care (intervention) with the following comparator strategies:

- ordinary referral to secondary care without prior primary care investigations for all patients ('refer all' strategy)
- primary care investigations alone
- send home/wait ('send home and wait' strategy).

In *Figure 5*, we illustrate in the diagnosis pathway for patients with low-risk symptoms who do not fulfil the 2WW criteria.

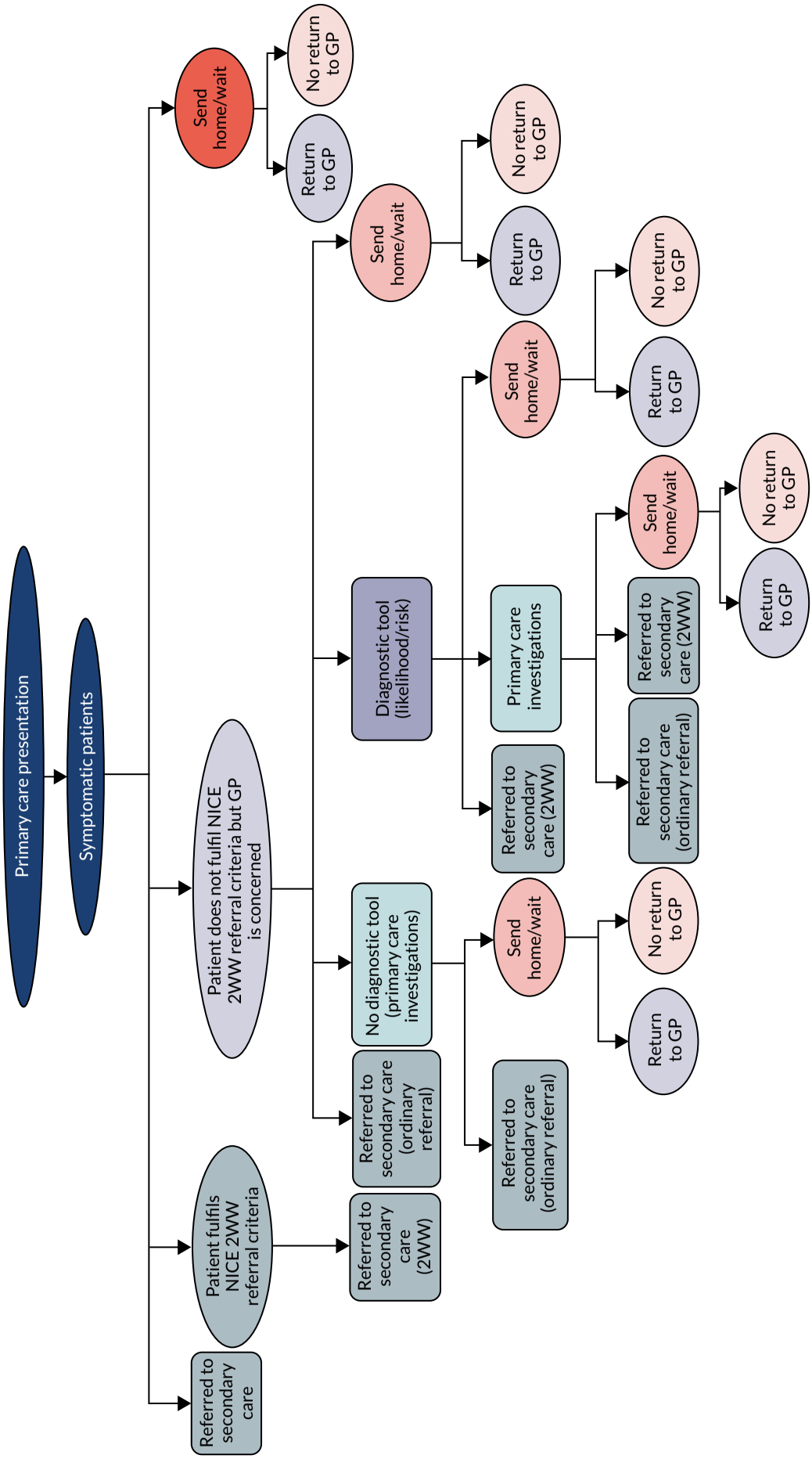


FIGURE 5 Diagnosis pathway model.

In the absence of clinical studies providing evidence on the clinical outcomes of these strategies, the effect of the diagnostic algorithm on patient management is modelled using a linked-data approach, which translates evidence of diagnostic accuracy into clinical outcomes in terms of life expectancy, health-related quality of life and health-care costs. In the model, the gain in sensitivity with the algorithm may cause an increase in the number of patients being referred to secondary care, which, in turn, results in the earlier identification of CRC cases; this is the main driver of outcomes in the model. An increase in sensitivity may also change the composition of diagnosis by disease stage, but this is harder to model without further data. The clinical benefits of the diagnostic tools depend on detecting cases at earlier disease stages than those that would occur under the status quo, without use of the diagnostic tools.

Population

To implement our model-based analysis, we define the relevant 'low-risk' population as those symptomatic patients presenting for the first time to a GP to seek care for their symptoms. We use prevalence estimates from Hippsley-Cox and Coupland,[80] who obtained data on the number of identified CRC cases from the electronic records of 224 primary care practices in the UK over a 2-year period (from January 2000 to April 2002) following the first visit of a patient with eligible symptoms to a GP. When patients in the top 1% risk stratum are excluded, the estimated CRC prevalence rate in this population is 0.16%.[80] As the mean age in the sample was 50.1 years, and prevalence reportedly increased 'steeply' with age, for the case of our modelled base-case analysis, aged 70 years, a prevalence rate of 1.5% is used, which is consistent with the age effects in the risk calculator published by the same authors[80] and is consistent with previous analyses informing policy.[140] No other sources of reliable prevalence data were identified.

This patient population is akin to the 'low risk but not no-risk symptoms'[75] population of the CAPER study.[75] The patient population includes the majority of symptomatic patients who do not present with high-risk symptoms, such as rectal bleeding or severe anaemia, and, therefore, are not diagnosed through the 2WW referral pathway. These are the populations the tools were developed for.

Interventions

The intervention involves use of the diagnostic tool and referral of those individuals above the threshold score of 35, which corresponds, approximately, to 2% in RAT and 0.5% in QCancer, to secondary care for investigation, which, in the model, is assumed to be colonoscopy. Although some patients undergo alternative tests such as flexible sigmoidoscopy, CTC and barium enema because of the presence of certain symptoms, older age or patient choice, only colonoscopy was modelled, as it is the most common test used.[155] This assumption was intended to avoid unnecessary complexity, given the exploratory aims of the model. For those individuals with a score below the threshold, the diagnostic tool is followed by primary care investigations, which, in the model, are assumed to consist of FITs at the 20-μg threshold (hereafter referred to as FIT). A subgroup of patients may also be sent home without a FIT if the presenting symptoms result in a score below a certain minimum level. We do not have adequate information on diagnostic accuracy for this group of patients deemed to be at too low a risk to warrant investigation; therefore, it is assumed that all patients below the threshold undergo a FIT. As a result, our analysis assumes that using the tools will always result in some further action. *Figure 6* shows the decision model for patients with CRC and *Figure 7* shows the decision model for patients without CRC.

A key difference between the diagnostic model for CRC patients and the model for non-CRC patients in *Figures 6* and *7* is that the proportion of patients referred to secondary care is based on the sensitivity (p1 and p2) of the tests for CRC patients and specificity (p3 and p4) of the tests for patients without CRC. In addition, patients without CRC are assumed not to return to general practice after a negative test result.

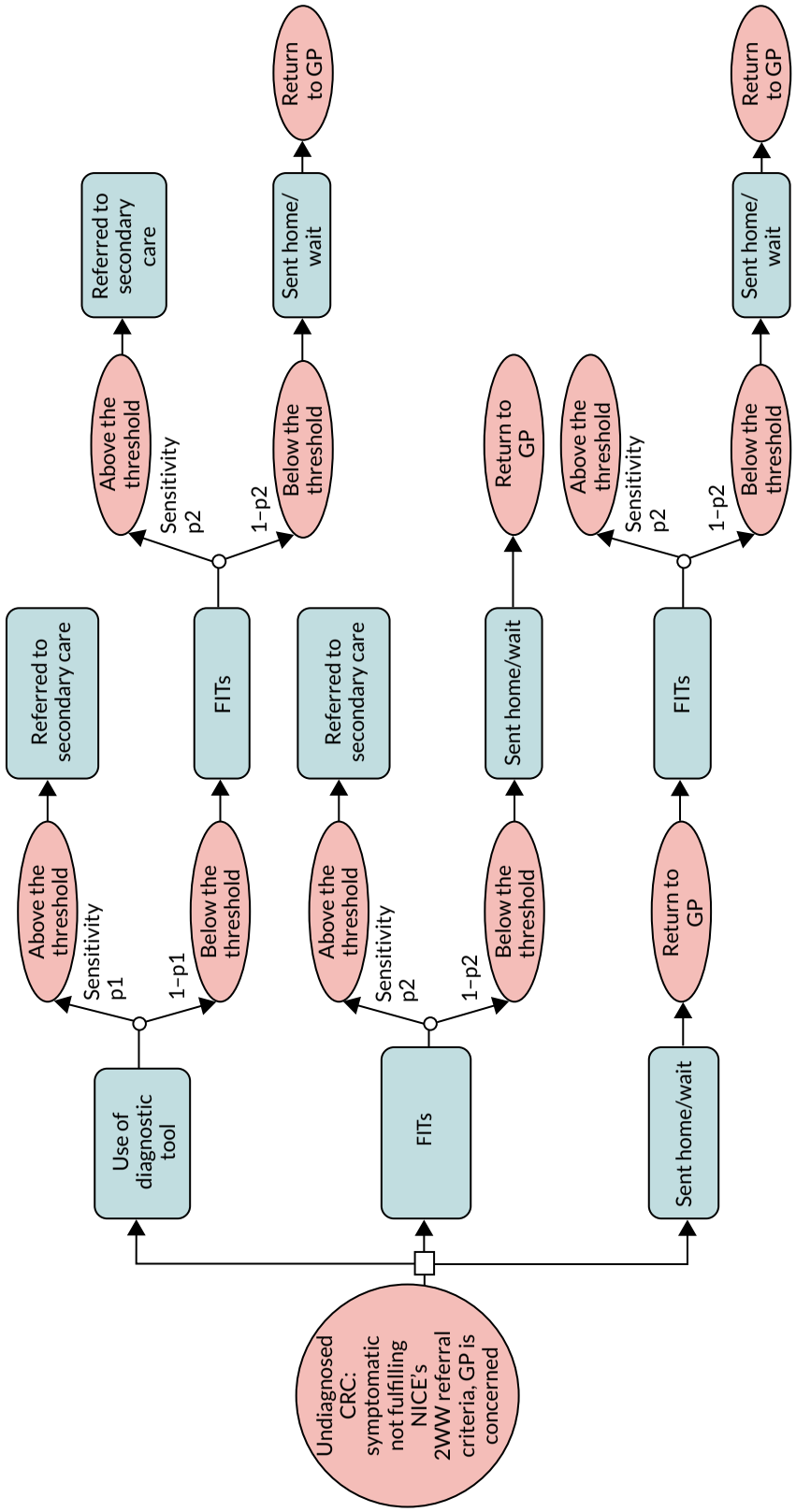


FIGURE 6 Diagnostic decision model for patients with CRC. p1, number of cases with a positive test result using the diagnostic tool/total number of CRC cases; p2, number of cases with a positive test result using the FIT test/total number of CRC cases.

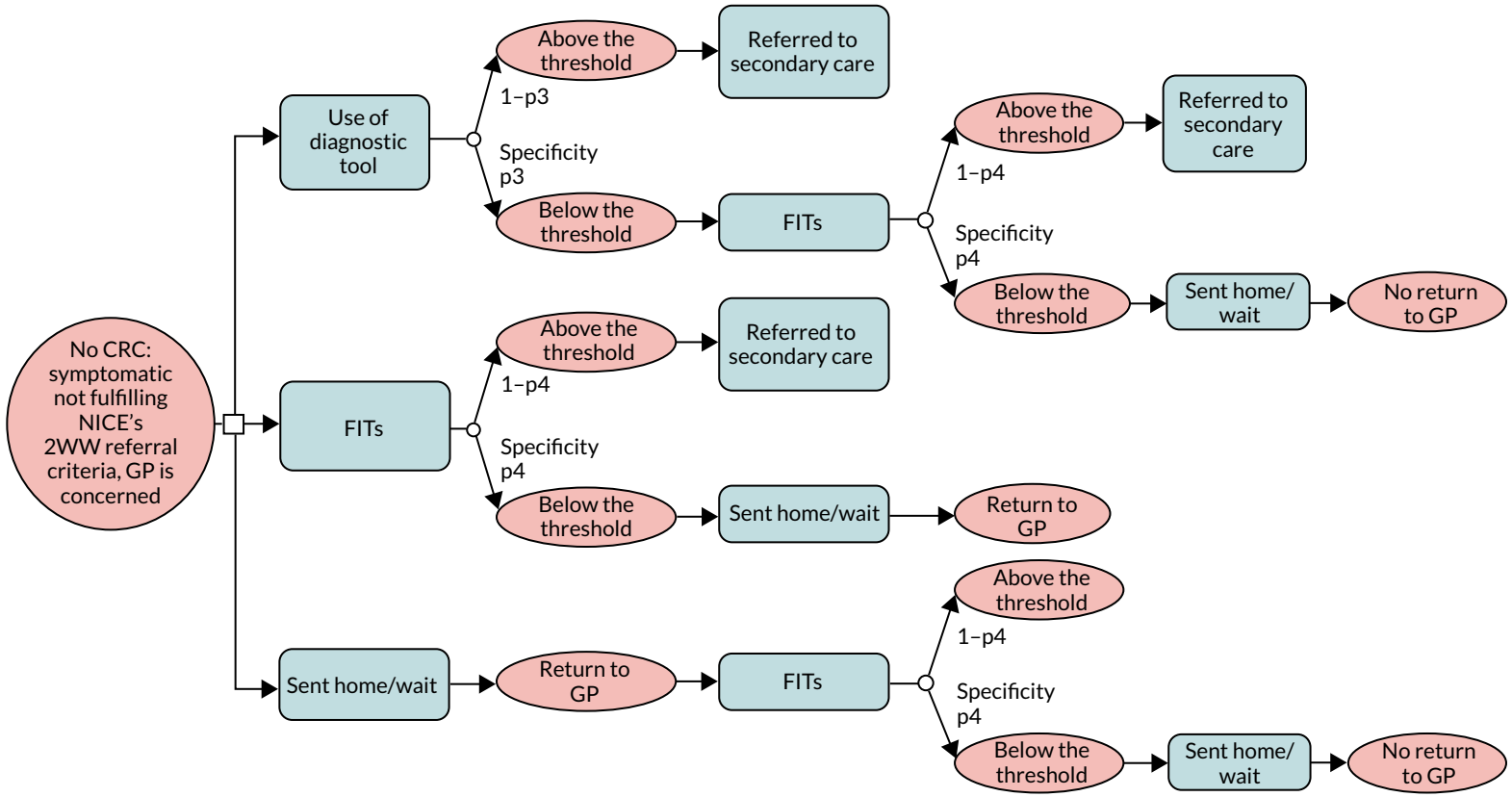


FIGURE 7 Diagnostic decision model for patients without CRC. p3, number of patients with a negative test result using the diagnostic tool/total number of patients without CRC; p4, number of patients with a negative test result using the FIT test/total number of patients without CRC.

Comparators

Faecal immunochemical test given to all

Based on clinical judgement, the standard diagnostic practice involves referral to secondary care after conducting FITs for all patients. Pragmatically, we assume that the referral decision rule consists of referring those patients who would have a positive FIT result, and sending home patients with negative results for watchful waiting. We chose to use this as the relevant comparator based on advice from local practitioners and current NICE Diagnostic Guideline recommendations,[140] and guided by the forthcoming roll-out of this test as a screening test nationally. Thus, we acknowledge that at the time this report is published FIT is not yet universally used.

Send home/wait

General practitioners may send a patient home without undertaking any primary care investigations, and wait until the symptoms fail to subside and the patient makes a repeat visit before ordering such investigations. We have evaluated this option as a scenario involving an initial visit whereby the patient is sent home without having undergone primary care investigation and a return visit the following month where all (undiagnosed) CRC cases and none of the non-CRC patients return for a FIT. We vary the proportion of non-CRC patients who return for a second visit in sensitivity analyses. All outcomes and patient management pathway from the second visit are as in the 'FIT given to all' pathway.

After the referral decision, those with correct negative results, the true negatives, are assumed not to return to the general practice. The same applies for those with incorrect positive results, the false positives. As for CRC patients with incorrect negative test results, the 'false-negative' cases, it is assumed that they will return 1 month later with persistent symptoms. The pathway for these patients and those correctly identified, the 'true-positive' cases, is further determined by the disease model described further on.

Refer all

A diagnostic strategy may be to refer all low-risk patients to secondary care. Given the low prevalence of CRC in this low-risk population,[80] between 0.1% and 3%,[140] this strategy is unaffordable and not used in practice; therefore, it is not considered further.[140]

Diagnostic pathway model

To predict the diagnostic interval associated with each of the diagnostic strategies, we made the following assumptions:

1. In the strategies involving testing, namely the two intervention strategies involving a tool plus FIT and the FIT-alone investigation strategy, patient management is determined by the result of the diagnostic test. In our base-case analysis, we allow for partial adherence to the diagnostic protocol, whereby 20% of positive results do not lead to referral or the referral is not realised.
2. In the intervention strategies whereby the tool produces a score below the threshold so that there is a sequence of two tests, the tool first and FIT second, we assume that the sensitivity and specificity of the second test in the sequence is independent of the outcome of the first test. This is likely to be an optimistic assumption, as discussed further on.
3. A CRC patient who remains undiagnosed after a first symptomatic presentation to a GP will have repeat monthly GP visits until diagnosis or death. In scenario analyses, we limit the number of repeat monthly visits to a maximum of 12, after which the GP refers the patient to secondary care without any further testing.
4. The sensitivity of the overall strategies determines the number of visits and monthly cycles before referral in the model. This assumption is what ultimately drives the clinical effectiveness in the model, which may be justified by the evidence available from Lyratzopoulos *et al.*[173] showing that the number of GP visits before diagnosis of CRC patients is positively associated with the median time to referral.

5. In the absence of data, we assume that the accuracy of diagnostic tests is independent of the disease stage at presentation. Therefore, the same intervention effect on the diagnostic interval relative to the comparators arising from assumptions 1, 2 and 3 leads to different clinical outcomes across initial disease stages, given their different rates of disease progression and responsiveness to treatment (see *Disease history model linking test results and clinical outcomes*). We obtain an aggregate predicted outcome for an initial distribution of disease stages in a simulated incidental cohort from our replication of the model described by Tappenden *et al.*[145]

Disease history model linking test results and clinical outcomes

To model the implications of avoidable delays in referral and diagnosis, we employed the updated version of the disease model developed by Tappenden *et al.*,[145] as described in Whyte *et al.*[156] As we could not obtain the model from Whyte *et al.*,[156] we replicated it from information published on its methods.[145,156,174,175] As this disease history model was not developed for evaluating diagnostic strategies for identifying cancer patients in primary care, we adapted it to focus entirely on the symptomatic stage population of present interest. An illustration of the adapted Markov model of health-state transitions experienced by CRC patients visiting their GP with symptomatic disease is presented in *Figure 8*. Briefly, an undiagnosed patient presents to primary care with symptoms at one of the four possible disease stages and is managed according to one of the diagnostic strategies investigated in this study. The patient's disease is detected and he/she is referred to secondary care for investigation with a certain probability (p), which we assume is equal to the sensitivity of the strategy, or is missed with a probability of $1 - p$. Once in secondary care, the patient undergoes colonoscopy and, because this investigation has sensitivity approaching 100%, the transition probability to a cancer diagnosis from being undiagnosed equals the overall sensitivity of the strategy, p, times the probability of surviving until the end of the cycle, $1 - d$.

Colorectal cancer patients missed by the diagnostic strategy remain untreated and are at risk of progressing to the next stage of disease severity. The probability of disease progression from one cycle to the next is equal to the probability of a false-negative result with the diagnostic strategy times the probability of remaining alive by the end of the cycle times the probability of disease progression (q), $q \times (1 - d) \times (1 - p)$. If, after repeated cycles of visits to the GP, the patient remains undiagnosed, the disease stage D would be reached, for which the only possible transition, other than to starting treatment, is to death from CRC or other causes.

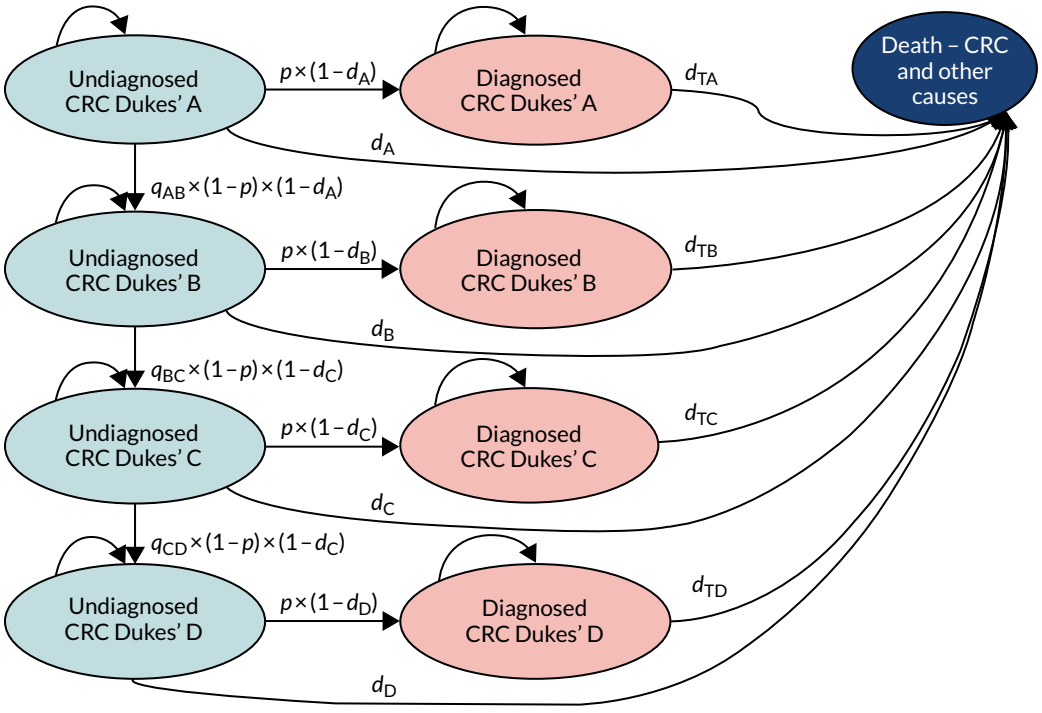


FIGURE 8 Markov model of health-state transitions for CRC cases.

In this model, a patient with CRC who, under the intervention, is identified earlier than under the control, say Dukes' stage A or B instead of C or D, derives both short-term and long-term health benefits. The short-term benefits occur as soon as treatment starts in the form of death risk reduction, whereas the longer-term benefits arise from avoiding disease progression to Dukes' stages C or D, for which treatment options are less effective. In terms of *Figure 8*, the excess death risk of not receiving treatment is captured by the relative magnitude of death probabilities on and off treatment for each of the Dukes' disease stages A, B, C and D: $d_{TA} < d_A$, $d_{TB} < d_B$, $d_{TC} < d_C$ and $d_{TD} < d_D$, respectively. The long-term survival loss from delays in diagnosis is measured by the positive relationship between the probability of CRC-related death and disease severity, that is $d_{TA} < d_{TB} < d_{TC} < d_{TD}$. The reduction in death risk is assumed to continue for the rest of the patient's life, apart from the maximum survival ceiling imposed by the background mortality risks included in the model, as measured in life tables[176] for the general English population of the same age and sex. In the scenario analysis, we explore the effects of adopting the assumption in Whyte *et al.*,[152] whereby patients who remain alive at the end of the 5-year period after diagnosis have, from then on, the same cycle probability of death as that of the general population of the same age and sex, that is negligible CRC death risk.

For patients with no CRC, a two-state Markov model was used that followed the initial decision tree of *Figure 5*. After the initial decision of whether or not to refer those patients, those referred undergo colonoscopy and are exposed to its associated small risk of adverse events, whereas those not referred are spared such exposure. Patients who are alive after colonoscopy or not referred to colonoscopy are in the alive health state and may transition to death from non-CRC causes, as determined by the death risks in English life tables, or remain alive at the start of the next cycle.

Mechanism of effect

In the absence of direct clinical effectiveness data, we assume a relationship between the sensitivity of both the algorithms and routine referral practice and the time to diagnosis. We break the time from symptom presentation to diagnosis into (1) the time from presentation to primary care to referral for investigation (the 'referral interval' in *Chapter 5*) and (2) the time from referral to specialist care to diagnosis and treatment, and assume that the diagnostic algorithms affect only the referral interval, while the time from referral to diagnosis and treatment remains fixed. Thus, in our analysis, a reduction in the diagnostic interval associated with any improvement in primary care diagnostic accuracy is equal to the reduction in the time from presentation to referral that is mediated through the number of primary care visits before referral.

Given our chosen population, we model the clinical effectiveness as a function of the diagnostic outcome of the first visit, assuming that, under current clinical practice, true-positive cases are referred after two visits, to allow for primary care investigations to be conducted and discussed between the patient and the doctor. False-negative cases would see their symptoms persist and would return for a third, or possibly more, visits. Because we have no obvious way to infer how a change in primary care diagnostic accuracy would affect the distribution of second and subsequent visits, we assume that the sensitivity and specificity are independent across multiple visits to primary care. This may be incorrect because individuals who are missed on an initial visit are more likely to be missed on a second visit, and so on. Indeed, we infer the probable extent of this in primary care referral practice in *Table 15* and investigate the implications in sensitivity analyses (see *Scenario analyses*). Thus, a probability of referral after two, three, four and five or more visits is calculated, respectively, as:

$$P(n = 2) = sensitivity. \quad (1)$$

$$P(n = 3) = (1 - sensitivity) \times sensitivity. \quad (2)$$

$$P(n = 4) = (1 - sensitivity)^2 \times sensitivity. \quad (3)$$

$$P(n \geq 5) = (1 - sensitivity)^3. \quad (4)$$

TABLE 15 Implicit average median time to referral under current standard practice

Referral after	Actual frequency[a]	Median days to referral[a]	Implicit sensitivity by visit[b]	Predicted frequency[c]	Median days to referral[d]	Source
2 visits	0.76	18	0.76	0.76[d]	18	Analysis of data from the English National Audit of Cancer Diagnosis in Primary Care 2009–10, on 1170 general practices (≈ 14% of all English practices)[173]
3 visits	0.12	39	0.50	0.18	39	
4 visits	0.05	56	0.42	0.045	56	
≥ 5 visits	0.06	122	1.0	0.01[d]	122	
Weighted average	1	29	N/A	1.0	25	

N/A, not applicable.
a Lyratzopoulos *et al.*[173]
b Weighted average calculated by the authors from data reported by Lyratzopoulos *et al.*[173]
c Assuming independence of sensitivity values across multiple visits.
d Weighted average calculated by the authors from data reported by Lyratzopoulos *et al.*[173] and predicted frequency distribution of number of pre-referral visits.

Under the diagnostic sensitivity of 0.76 implicit in the results reported by Lyratzopoulos *et al.*,[173] the predicted frequency for three, four and five or more visits is 18%, 4% and 1%, respectively, which is in contrast with the actual frequencies of 12%, 5% and 6%, respectively. Consequently, the predicted median referral interval of 25 days is lower than the actual figure of 29 days (see *Table 15*).

The length of time to referral (i.e. diagnosis in our model) is derived from the number of primary care visits from previous research.[173] Time to referral and the number of primary care visits before referral are strongly positively associated, with a threshold of 16 or 17 days discriminating between cancer patients referred after three or more and two or fewer visits.[173]

In our Markov model of 28-day cycles, the length of the referral interval is equal to the number of initial repeat GP visits times 28 days (*Table 16*). The resulting times to referral implicit in the sensitivity values of the tests in each strategy adopted for the model are presented in *Table 16*. These base-case values correspond to sensitivity and specificity values of 69% and 77%,[141] 61% and 91%[80] and 53% and 99%[175] for the RAT, QCancer and FIT of 20 µg Hb/g faeces, respectively, after adjusting for the 80% compliance with diagnostic results, so that 20% of those testing positive would not be referred but sent home. These values are an approximation to the model results, which also account for the attrition caused by death from CRC or other causes before referral, so that the values in *Table 16* are greater than the model outputs by 2–2.5 days. When the sensitivity values implicit in the number of pre-referral visits reported in a retrospective patient experience survey (Lyratzopoulos *et al.*[173]) are used for FITs, the reduction in the time to referral relative to the comparator is only 7 days for QCancer and 8 days for the RAT (see *Appendix 5*).

A note of warning is due as to the lack of comparability between the accuracies of the RAT and QCancer. Their specificity and sensitivity values were derived from studies using different designs (a case–control study for the RAT and cohort data from electronic records for QCancer) and populations. This would render any comparison of the tools biased. We note that similar arguments could also be raised in relation to the comparison of each tool with the comparator, as this is also based on an indirect comparison of heterogeneous single-arm accuracy studies; this fact reflects the low quality of data available for analysis.

To complete the link from diagnostic accuracy outcomes and number of cycles before referral (which determines the time to treatment because the time from referral to diagnosis and treatment is assumed to be unaffected) with survival outcomes, we use the relationship between time to diagnosis

TABLE 16 Mean time to referral of CRC patients by strategy in the model: base-case analysis

	QCancer[a]		Comparator[b]			RAT[a]		Comparator[b]	
Referral after	Predicted frequency[c]	Mean time to referral (days)	Predicted frequency[d]	Mean time to referral (days)	Referral after	Predicted frequency[c]	Mean time to referral (days)	Predicted frequency[d]	Mean time to referral (days)
1 visit	0.49	28	N/A	28	1 visit	0.55	28	N/A	28
2 visits	0.27	28	0.42	28	2 visits	0.21	28	0.42	28
3 visits	0.12	56	N/A	N/A	3 visits	0.13	56	N/A	N/A
4 visits	0.07	56	0.24	56	4 visits	0.05	56	0.24	56
≥ 5 visits	0.06	93	0.34	123	≥ 5 visits	0.06	93	0.34	123
Weighted average	1	37	1	67	Weighted average	1	37	1	67

N/A, not applicable.
a Derived from the sensitivity values for the component tests times the 80% rate of compliance of referral practice with the test results; see *Table 19*.
b The modelled comparator strategy assumes that every time the patient returns two GP visits are required for a referral decision to be made, one for the return visit and second, follow-up visit, for the review and discussion of the results of primary care investigations ordered in return visit.
c Positive test results with the tool result in referral after the first visit, whereas positive test results with a FIT after negative result with the tool takes two visits.
d Positive test result with FIT alone results in referral after two visits, an initial and a follow-up visit to discuss results.

and disease stage at diagnosis. The only available source for this unobservable parameter is provided by a calibration exercise of a decision model based on Tappenden *et al.*[155] It reports rates of progressive transition between (non-clinical) Dukes' stages A to B, B to C and C to D, and from each of these to death, which, for stages earlier than D, was exclusively unrelated to cancer. We assume that the disease progression rates from this source, which were intended to depict the experience of patients in asymptomatic states, apply to our symptomatic population, but combine these with CRC mortality rates reported by disease stage in untreated patients (Liu *et al.*[177]) *Table 17* presents the mean time to progression from Dukes' stages A to B, B to C and C to D implied by the disease progression rates adopted.

The reductions in time to diagnosis and treatment made possible by the smaller number of initial repeat visits to the GP have the ultimate impact of increasing the likelihood of starting treatment at a stage of disease for which treatment is more effective. We employ data on relative survival (with respect to the general population) by disease stage at diagnosis, depicted in *Figure 9*, to model this longer-term health benefit of earlier access to treatment with more accurate GP diagnosis and referral.[178] We present the expected survival over time of an individual who is diagnosed while at Dukes' stage B and the counterfactual, improved survival, scenario that would have been observed had this simulated case been diagnosed earlier while still in Dukes' stage A (*Figure 10*).

TABLE 17 Dukes' stage transitions

Dukes' stage transition	Annual transition probability (%)	Mean time to progression (months)	Source
A to B	58.29	12.3	Authors' calculations from reported model-calibrated data[a]
B to C	65.55	9.0	
C to D	86.48	4.3	

a Tappenden *et al.*[145]

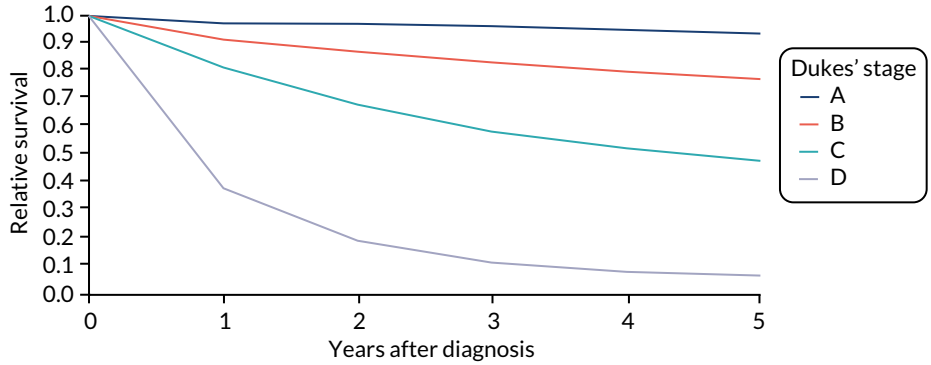


FIGURE 9 Relative survival of CRC patients diagnosed in 1996–2002 in England. Survival curves by disease stage. Source: National Cancer Registration and Analysis Service.[178]

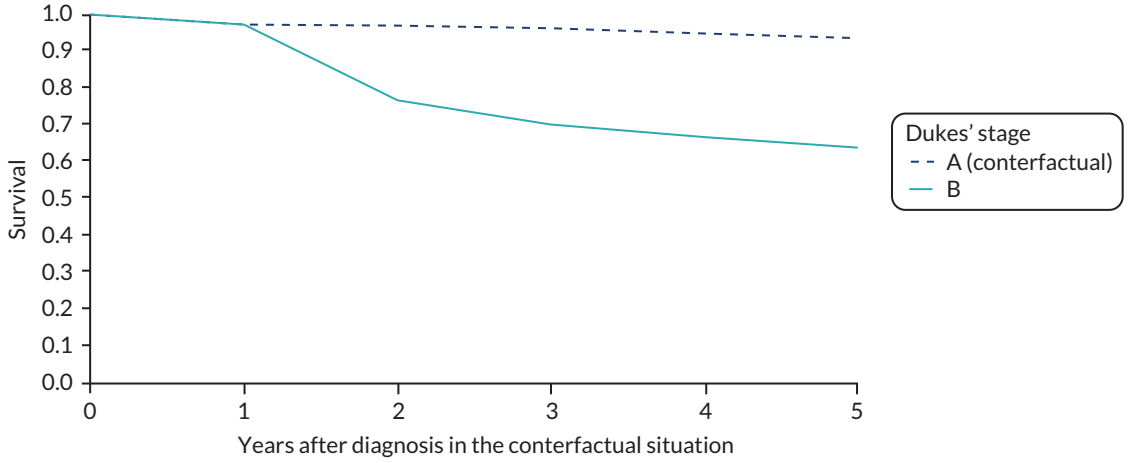


FIGURE 10 Actual and counterfactual survival curves for a patient who is diagnosed at Dukes' stage B.

In *Table 18*, the link between diagnostic accuracy and disease stage at diagnosis, which is mediated by the number of pre-referral visits (cycles) and underpins the clinical effectiveness in the model, is made explicit for each strategy. The more accurate intervention strategies shift the distribution of disease stage at diagnosis towards the earlier disease stages of Dukes' stages A and B, and C (not shown in the table). The overall percentage of stages A/B presented in the table do not account for background mortality effects;

TABLE 18 Disease stage at diagnosis of CRC patients by strategy in the model: base-case analysis

	QCancer		Comparator			RAT		Comparator	
Referral after	Predicted frequency	Dukes' stage A/B (%)	Predicted frequency	Dukes' stage A/B (%)	Referral after	Predicted frequency	Dukes' stage A/B (%)	Predicted frequency	Dukes' stage A/B (%)
1/2 visits	0.65	54	0.42	54	1/2 visits	0.68	54	0.42	54
3/4 visits	0.23	51	0.24	51	3/4 visits	0.22	51	0.24	51
5/6 visits	0.08	48	0.14	48	5/6 visits	0.07	48	0.14	48
7/8 visits	0.03	45	0.08	45	7/8 visits	0.02	45	0.08	45
9/10 visits	0.01	42	0.05	42	9/10 visits	0.01	42	0.05	42
≥ 10 visits	0.01	≤ 40	0.07	≤ 40	≥ 10 visits	0.00	≤ 40	0.07	≤ 40
Weighted average	1	52	1	50	Weighted average	1	53	1	50

therefore, they differ by about 1 percentage point from the final model estimates, which do account for such mortality. When the sensitivity values implicit in the number of pre-referral visits reported in a retrospective patient experience survey (i.e. Lyratzopoulos *et al.*[173]) are used for FITs, there is a 1-percentage-point gain in the proportion of patients with Dukes' stages A/B at diagnosis with both QCancer and the RAT, relative to the comparator (see *Appendix 5*). Therefore, an improvement in the sensitivity of primary care diagnostic testing would be expected to reduce the time to referral and, provided this is translated into a one-to-one reduction in the time to diagnosis and treatment, increase the relative frequency of earlier disease stages at diagnosis. Stage at diagnosis would then determine the long-term clinical outcomes and costs in the model.

The values of key model parameters are presented in *Table 19*. These values were identified from reviews of previous decision models in CRC, described in *Chapter 6*. This choice of data sources was partly determined by the fact that, in contrast to our research plans, the decision model of Whyte *et al.*[156] was not made available to our study. As a result, we had to reallocate the research resources that had been originally assigned for conducting a systematic review of model parameter values to building the model from scratch by replicating its published methods. Further details of the methods and data used to populate the disease model parameters are given in *Appendix 5*.

TABLE 19 Model parameter values used in the base-case analysis

Parameter	Value	Comment	Source
Clinical			
Starting age, 70 years			
Prevalence of CRC	0.015	Symptomatic low-risk population, as in analysis informing NICE Diagnostic Guideline Number 30[140]	Hippsley-Cox 2012[80]
Disease stage at presentation		Age-specific distribution by Dukes' stage	Simulated by authors from replicated disease history model (Tappenden 2007[145])
Sensitivity of QCancer	0.610	Based on the top 10% risk score in their tool, which gives a risk threshold of 0.5% (PPV 1.5%). We recalculated their sensitivity and specificity of this threshold to exclude observations in the top 1% risk score range (5.2% risk threshold) and thus get closer to the low-risk population (0.1–3.0% prevalence); unfortunately, the source does not give the required breakdown of results to exclude people above the 3% threshold	Hippsley-Cox 2012[80]
Specificity of QCancer	0.910		
Sensitivity of the RAT	0.69	Based on the 2% risk threshold	Hamilton 2005[141]
Specificity of the RAT	0.77		
Sensitivity of the FIT	0.526	FIT 20 µg Hb/g faeces	Murphy 2017[175]
Specificity of the FIT: 50–69 years old	0.988		
Specificity FIT: ≥ 70 years old	0.963		
Compliance with colonoscopy referral	0.80		Tappenden 2007[145]
Colonoscopy sensitivity for CRC	1	Murphy 2017[175] reports a value of 0.966 from a systematic review (i.e. van Rijn 2006[179])	Assumption
Colonoscopy specificity for CRC	1	Assumption	Murphy 2017[175]
Probability of perforation with colonoscopy	0.0017		Atkin 2002[180]
Probability of death with perforation	0.0582		Gatto 2003[181]

continued

TABLE 19 Model parameter values used in the base-case analysis (*continued*)

Parameter	Value	Comment	Source
Probability of bleeding	0.0044		Atkin 2002[180]
Transition probabilities		4 weekly	Whyte 2014[156]
Dukes' stage A to B	0.0651	Reported model calibration to CRC incidence and stage at diagnosis	Tappenden 2007[145]
Dukes' stage B to C	0.0787		
Dukes' stage C to D	0.1427		
CRC death risk Dukes' stage A	0.00	Post Dukes' stage A at diagnosis	Exponential hazard function fitted to digitised Kaplan–Meier curves up to 5-year relative survival for CRC patients (diagnosed 1996–2002) by stage at diagnosis in England[178]
CRC death risk Dukes' stage B	0.00	Post Dukes' stage B at diagnosis	
CRC death risk Dukes' stage C	0.01	Post Dukes' stage C diagnosis	
CRC death risk Dukes' stage D	0.07	Post Dukes' stage D at diagnosis	
HR untreated CRC in Dukes' stage A	1	Survival of patients refusing treatment with newly diagnosed CRC in Taiwan (Province of China) during 2004–8. Treatment refusal was defined as 'not undergoing any cancer treatment within 4 months of confirmed cancer diagnosis[177]'	Liu 2014[177]
HR untreated CRC in Dukes' stage A	1.22		
HR untreated CRC in Dukes' stage A	1.22		
HR untreated CRC in Dukes' stage A	1.22		
Costs			
Cost of FIT	£3	Threshold 20 µg Hb/g faeces. Published data, excludes costs of screening campaign in the original source	Murphy 2017[175]
Cost of GP visit	£31	PSSRU's unit cost per surgery consultation lasting 9.22 minutes, without qualification costs and including direct care staff costs	Curtis 2018[182]
Cost of colonoscopy	£615	Costs of colonoscopy with or without polypectomy	Whyte 2014,[156] NHS reference costs, screening centre estimates[152]
Cost of treating bowel perforation	£5559	Major surgery	Whyte 2014,[156] NHS reference costs
Cost of bleeding	£304	Cost of admittance for bleeding (overnight stay on medical ward), NHS reference costs	Whyte 2014[156]
Health-care cost, Dukes' stage A	£3178	• Lifetime estimate, age specific, figure presented is for people aged 70–79 years • Published data from whole-disease model simulations	Whyte 2012[174]
Health-care cost, Dukes' stage B	£3455		
Health-care cost, Dukes' stage C	£4485		
Health-care cost, Dukes' stage D	£4365		
Utilities			
Cancer free	0.91	Whyte 2014.[156] General population utility decline with age by sex was accounted for using the linear model proposed by Ara 2010[183]	Ness 1999[184]
Dukes' stage A	0.74		
Dukes' stage B	0.70		
Dukes' stage C	0.50		
Dukes' stage D	0.25		

HR, hazard ratio; PSSRU, Personal Social Services Research Unit.

Calculations of costs and benefits in the model

Our Markov model consists of a repeated iteration of 28-day cycles, whereby patients accrue costs and benefits depending on the health state occupied in each iteration. We chose this cycle length as the most appropriate for capturing the frequency with which patients with unresolved symptoms return to their GPs for care. By aggregating the proportion of patients occupying the various health states in each 28-day cycle across the modelled time horizon of 30 years, we calculated the total time alive and total time spent in each state. We applied costs and health-state utility values corresponding to the different states and aggregated across health states and time periods to calculate total health-care costs and total QALYs. The exception to this method is the estimation of health-care costs of cancer treatment, which are implemented as a single lifetime cost payoff incurred at diagnosis, as in the original model by Whyte *et al.*[156] The modelled accumulated total life-years, QALYs and lifetime health-care costs for the intervention and the status quo strategies, over a period of 30 years after an initial presentation with symptoms to the GP, are presented for a hypothetical cohort aged 70 years at first presentation. We initially focus on the predicted clinical outcomes for the subgroup of CRC cases and then present the results for the overall cohort presenting to their GP with symptoms suggestive of CRC.

Table 20 compares the main features of our model with existing models relevant to our study question. Common to ours and the previous models is the use of a disease history model developed by Tappenden *et al.*[145] Whyte *et al.*'s[156] model is a model of screening and therefore does not address the question of interest here. The main differences between our and Westwood *et al.*'s[161] work is that we explicitly model the time to referral consistent with the diagnostic accuracy of the respective test strategies and the distribution of the number of visits before referrals; we chose a shorter cycle length (28 days *vs.* annual) to allow greater granularity for predicting outcomes; and we model the short-term survival benefits of reducing the time to referral and treatment. Like Westwood *et al.*'s[161] and Whyte *et al.*'s[156] models, we have implemented a probabilistic sensitivity analysis, but with the caveat that this type of sensitivity analysis was of limited relevance because most of the uncertainty derives from structural assumptions, such as how the length of the diagnostic and treatment interval is determined, defining the status quo diagnostic practice and measuring its diagnostic accuracy. Consequently, we did not conduct any value-of-information analysis.

Deterministic sensitivity analysis

To investigate areas of uncertainty in our model, we arbitrarily varied clinical parameters by 20% above and 20% below their base-case analysis values. In addition, we undertook scenario analyses for the following parameters:

- alternative sensitivity and specificity values for the comparator (inferred from the number of GP visits before referral[173] and from receiver operating characteristic data for version 3 of NICE guidelines[22])
- assuming 100% compliance with the test result (i.e. all positive cases are referred and all negative cases are not referred to secondary care)
- no mortality benefit of treatment in stages A/B, by setting the hazard ratio to 1
- restricting excess CRC-related death risk to first 5 years[156]
- limit time horizon of analysis to 5 years
- alternative utility values adopting a common value for all CRC disease stages[145]
- alternative 'send home/wait' comparator strategy, whereby, at the initial visit, the GP sends the patient home without conducting a FIT; from the second visit onwards, they manage the patient as in the 'FIT given to all' strategy
- modified intervention strategy whereby patients who are above the risk threshold of the tool are referred directly to secondary care, as in the original intervention strategy (see *Figure 2*), whereas those below the threshold are sent home, rather than being subjected to primary care investigations, as assumed in our base-case analysis
- assuming no survival gains and maximum quality-of-life (utility) gains (i.e. zero utility values for false-negative cases until diagnosis) from more expeditious referral and diagnosis in patients with CRC.

TABLE 20 Key model characteristics of the de novo model and previous models

Study	Perspective	Population	Diagnostic strategies	Outcome measure(s)	Model structure	Model duration	Uncertainty analysis	Discount rate (%)	Base year and currency
Our current model	UK NHS	Symptomatic patients aged ≥ 40 years presenting to primary care who are at low risk of CRC	• QCancer + FIT • RAT + FIT • FIT alone	Cancer mortality, disease stage at diagnosis, life-years, QALYs	Markov model with monthly cycles	Lifetime	• Threshold analysis (diagnostic accuracy) • Scenario analysis	3.5	2018; GBP
Whyte 2014[156]	UK NHS	General population	N/A	Cancer mortality, QALYs	State-transition model with annual cycles	Lifetime	Two-way sensitivity and probabilistic sensitivity analyses	3.5	2012; GBP
Westwood 2017[161]	UK NHS	Symptomatic patients aged ≥ 40 years presenting to primary care who are at low risk of CRC	1. FIT 2. FOBT 3. No triage, refer all to colonoscopy	QALYs, life-years	Markov state-transition model with annual cycles	Lifetime	Scenario analysis and probabilistic sensitivity analysis	3.5	2015; GBP

GBP, Great British pounds; N/A, not applicable.

We present our results in 2017/18 prices. We provide tentative estimates of incremental costs per life-year gained. We also present results in terms of incremental cost per QALY gained, discounting costs and QALYs at an annual rate of 3.5%.

Results

Colorectal cancer patients

According to our model predictions, using the diagnostic tool triage interventions allow for sensitivity of 82–85%, compared with 53% for the current standard strategy of FIT for all.[175] The additional sensitivity of diagnosis would reduce the number of mean visits from 5.6 to 2.6. This, in turn, would lead to fewer patients dying undiagnosed (from 12% in the status quo to 8% with the diagnostic tools) and one additional case being diagnosed and treated in disease stages A/B for every 42 CRC patients [1/(0.009 + 0.015)] presenting to primary care with symptoms and suspected CRC (*Table 21*). The largest gains in life for a person aged 70 years occurs among individuals presenting with undiagnosed CRC of Dukes' stage B (0.26 years) and stage C (0.31 years), whereas the smallest gain occurs among those presenting with stage D CRC, who derive a gain of 14 days with either diagnostic tool.

Use of a diagnostic tool is predicted to reduce the amount of time spent in untreated health states by about 0.6 months (18 days), and to increase the amount of time alive while on treatment by 3 months, mostly after diagnosis at Dukes' stage A or B (*Table 22*). The net effect is to increase life expectancy by 2.4 months, at an extra cost of between £29 and £36. These figures do not account for costs and health outcomes in patients without CRC, which are included in the whole-population analysis next.

Whole-population analysis

When the costs and life-years are calculated over the denominator of all patients presenting with symptoms of suspected CRC, the importance of the trade-off between sensitivity and specificity becomes apparent. Detecting more CRC cases with diagnostic tool triage comes at the cost of imposing more unnecessary use of health-care resources and health risks from invasive investigations on the great majority of patients who do not have CRC. The costs associated with false-positive cases drives the cost difference between the diagnostic tool plus FIT and FIT alone, resulting in an extra cost of £42 or £105, depending on the diagnostic tool, per patient presenting with symptoms of suspected CRC. At the 1.5% prevalence rate of this low-risk population, the (undiscounted) additional cost per life-year saved is (42/0.0031) £13,548 with QCancer and (105/0.0032) £32,813 with the RAT (*Table 23*).

TABLE 21 Clinical model outcomes of each strategy for CRC patients aged 70 years vs. a FIT of 20 µg Hb/g faeces

Primary care diagnostic strategy	Overall sensitivity (%)	Mean number of GP visits	Dukes' stage at diagnosis				Percentage dying without being diagnosed	Life expectancy by Dukes' stage at first presentation				
			A	B	C	D		A	B	C	D	Weighted mean
FIT	53	5.6	16.9	32.3	24.0	14.7	12.1	13.52	9.29	4.47	0.95	7.22
QCancer + FIT	82	2.6	17.9	33.8	25.4	14.9	8.0	13.68	9.55	4.77	0.99	7.43
Difference: QCancer – FIT	N/A	6.4	0.9	1.5	1.4	0.0	–4.1	0.16	0.26	0.31	0.04	0.21
RAT + FIT	85	2.4	18.0	34.0	25.6	14.9	7.6	13.69	9.56	4.79	0.99	7.44
Difference: RAT – FIT	N/A	5.0	1.0	1.7	1.6	0.1	–4.5	0.17	0.28	0.33	0.04	0.23

N/A, not applicable.

TABLE 22 Economic model outcomes of each strategy for CRC patients aged ≥ 70 years vs. a FIT of 20 µg Hb/g faeces

Primary care diagnostic strategy	Mean number of GP visits	Mean time spent in each health state (months)								Life-years	QALYs	Cost (£)
		A	B	C	D	TA	TB	TC	TD			
FIT alone	5.6	0.2	0.4	0.3	0.2	29.1	39.0	15.2	2.2	7.22	3.65	4428
QCancer + FIT	2.6	0.1	0.2	0.1	0.1	30.3	40.4	15.8	2.2	7.43	3.87	4458
QCancer + FIT minus FIT alone	−3.0	−0.1	−0.2	−0.2	−0.1	1.2	1.4	0.6	0.0	0.21	0.22	29
RAT + FIT	2.4	0.1	0.1	0.1	0.1	30.3	40.5	15.9	2.2	7.44	3.89	4464
RAT + FIT minus FIT alone	−3.2	−0.1	−0.3	−0.2	−0.1	1.2	1.5	0.7	0.0	0.23	0.24	36

'TA', 'TB', 'TC' and 'TD' refer to Dukes' stages treated.

TABLE 23 Mean costs and QALYs per patient aged 70 years suspected of having CRC (0.18% prevalence) vs. a FIT of 20 µg Hb/g faeces

Outcome	QCancer + FIT	FIT alone	Difference: QCancer + FIT minus FIT alone	RAT + FIT	FIT alone	Difference: RAT + FIT minus FIT alone
Costs by initial test result (£)						
TN	58	62	−4	51	62	−11
FP	65	20	45	135	20	114
TP	44	29	16	46	29	18
FN	23	37	−14	21	37	−16
Total	190	148	42	252	148	105
Life-years						
TN	13.8023	14.8650	−1.0627	12.1507	14.8650	−2.7143
FP	1.5159	0.4534	1.0626	3.1674	0.4534	2.7140
TP	0.0732	0.0472	0.0260	0.0766	0.0472	0.0294
FN	0.0382	0.0610	−0.0228	0.0350	0.0610	−0.0260
Total	15.4297	15.4265	0.0031	15.4297	15.4265	0.0032
QALYs						
TN	10.1048	10.8827	−0.7780	8.8956	10.8827	−1.9872
FP	1.1098	0.3319	0.7779	2.3189	0.3319	1.9870
TP	0.0399	0.0257	0.0142	0.0417	0.0257	0.0160
FN	0.0182	0.0290	−0.0108	0.0166	0.0290	−0.0123
Total	11.2727	11.2694	0.0033	11.2729	11.2694	0.0035

FN, false negative; FP, false positive; TN, true negative; TP, true positive.

Colonoscopies contribute the largest source of difference in discounted costs, reflecting differences in specificity between strategies. The difference in colonoscopy costs relative to FIT alone is £42.34 for the QCancer plus FIT strategy, and £107.67 for the RAT plus FIT strategy (*Table 24*). Both intervention strategies have additional costs due to perforation and bleeding with colonoscopy, relative to FIT alone. Furthermore, there is an accompanying loss of QALYs of 0.0002 (equivalent to a loss of 1.75 full health-hours of life) associated with this unnecessary exposure to risks from colonoscopy in patients who do not have CRC, as a result of using the tools.

TABLE 24 Breakdown of discounted costs and QALYs in suspected cancer population

Costs and QALYs	FIT alone	QCancer + FIT	Difference	RAT + FIT	Difference
Costs pre diagnosis (£)					
GP visits	62.69	59.65	−3.04	56.17	−6.52
FITs	3.14	2.86	−0.28	2.51	−0.63
Colonoscopy	26.68	69.02	42.34	134.36	107.67
Adverse events with colonoscopy	0.47	1.21	0.74	2.36	1.89
Costs post diagnosis (£)					
Lifetime cost of treatment	54.84	56.83	1.99	56.97	2.13
Total	147.82	189.58	41.75	252.37	104.54
QALYs					
No CRC	8.8281	8.8281	−0.0001	8.8280	−0.0002
CRC pre diagnosis	0.0007	0.0003	−0.0004	0.0003	−0.0005
CRC post diagnosis	0.0482	0.0501	0.0019	0.0502	0.0021
Total	8.8770	8.8785	0.0015	8.8785	0.0015
ICER (£)			28,704		71,863

Tornado analysis

A tornado analysis is used to provide a graphical representation of the degree to which the incremental cost per QALY gained from tools is sensitive to independent variation by 20% of clinical parameters (*Figure 11*). The tornado analysis shows that the source of greatest uncertainty in the incremental cost per QALY originates from the sensitivity of the standard test strategy. A reduction of 20% in the false-negative rate, that is an increase in the sensitivity of the FIT from its base-case value of 53% to 62%, increases the discounted cost per QALY gained with the QCancer tool from £28,704 to somewhere above £48,633 (see *Figure 11*). A reduction of 20% in the same parameter reduces that figure to £17,842 per QALY gained. The algorithm's specificity is the second most important economic parameter, as it determines how much of the increased detection of CRC by the primary care diagnostic workup is obtained at the expense of patients without CRC who are referred, but have no need, for colonoscopy. Naturally, the cost of colonoscopy is the single most influential cost parameter in the results; reducing its cost from £615 to £492 reduces the cost per QALY gained with QCancer to £22,883.

Probabilistic analysis

A probabilistic sensitivity analysis was conducted for the base-case analysis by accounting for sampling uncertainty in the key model parameters (see the tornado plot in *Figure 11* and see *Appendix 5*, *Table 66*, for information on the distributions used for those parameters). The first and third quartiles of the incremental discounted costs' distribution were £40 and £44 per patient, respectively, whereas those for incremental discounted QALYs were 0.0012 and 0.0017, respectively (*Figure 12*). The probabilistic estimate of the incremental cost per QALY gained with the tool using the case of QCancer is £28,522, which is very similar to the deterministic estimate of £28,703.

Figure 13 presents the probability of the tool being cost-effective as a function of the willingness to pay for one QALY gained. This figure implies a 95% CI for the incremental cost per QALY gained with the tool of £17,000–50,000. A note of caution is warranted as this figure does not capture the uncertainty inherent in the assumption that the health benefits with the tool are due to its increased sensitivity, as it translates to fewer visits and, therefore, shorter times to referral, and that these reduced times in turn translate into diagnoses at earlier disease stages. Owing to this underestimation of the uncertainty in this probabilistic sensitivity analysis, we did not attempt to analyse the expected value of additional research information.

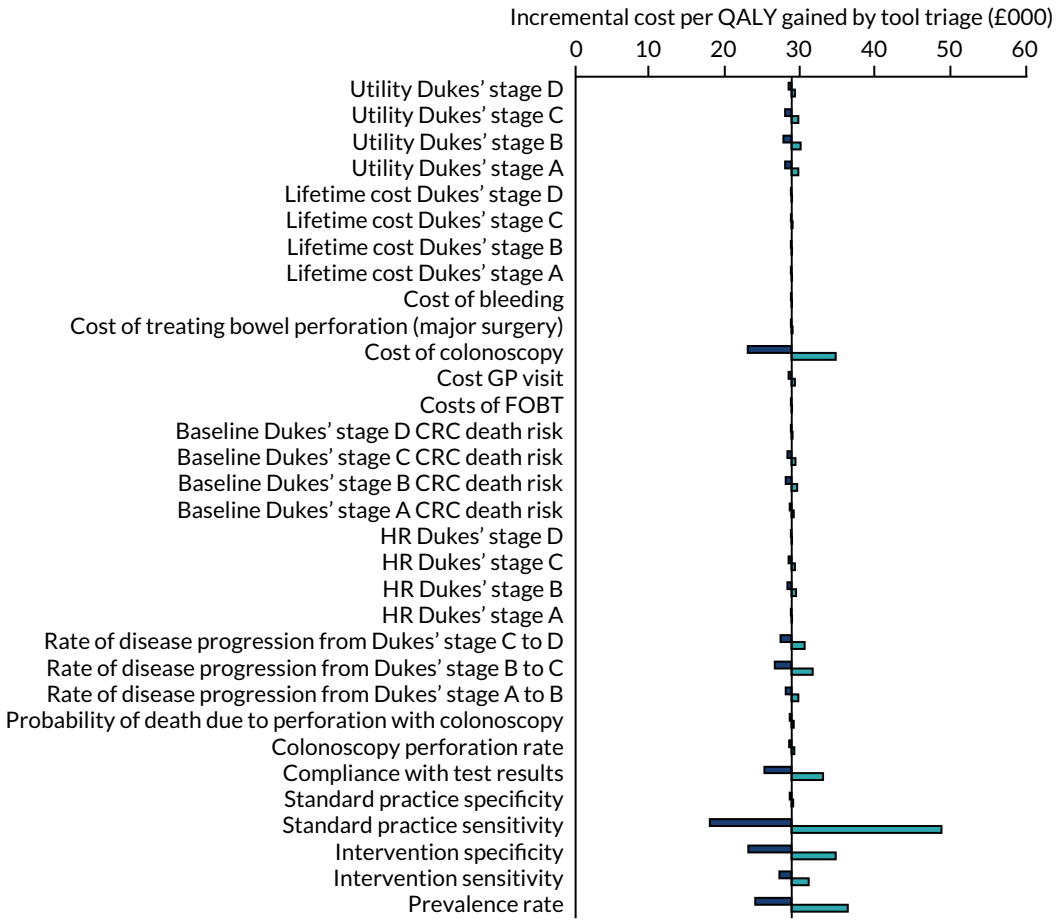


FIGURE 11 Sensitivity of results to arbitrary increases and decreases of 20% in model parameter values.

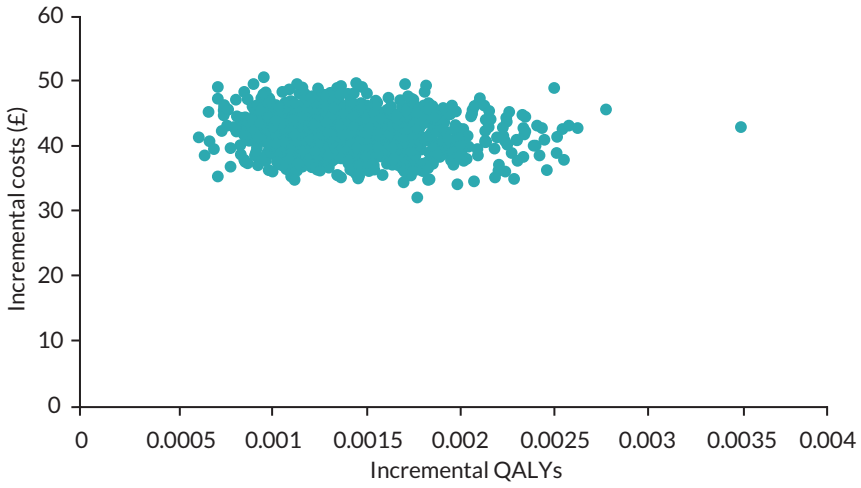


FIGURE 12 Cost-effectiveness plane: diagnostic tool (QCancer) vs. no tool.

Threshold analysis

The tornado analysis has highlighted the sensitivity of the cost per QALY gained to the sensitivity of the FITs and the specificity of the diagnostic tools. Given the key role of these parameters in the results, *Figure 14* depicts the levels of specificity of the triage tool required for the intervention to meet the £30,000 threshold of cost-effectiveness, at varying levels of sensitivity of the FIT. At the levels of sensitivity of ≥ 50% that have been reported for a FIT of 20 µg Hb/g faeces,[175] the minimum level of specificity needed by the tool to serve as a cost-effective triage intervention in primary care is 0.89.

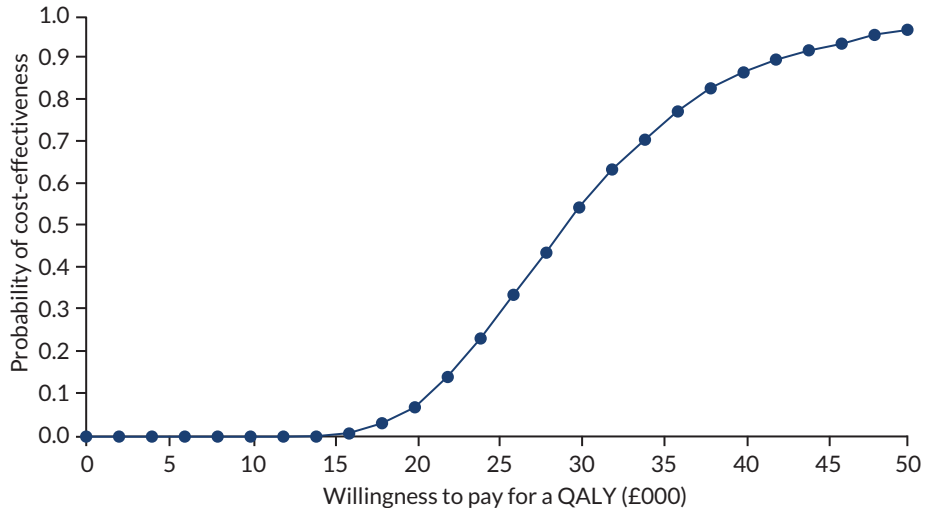


FIGURE 13 Cost-effectiveness acceptability curve: tool (QCancer) vs. no tool.

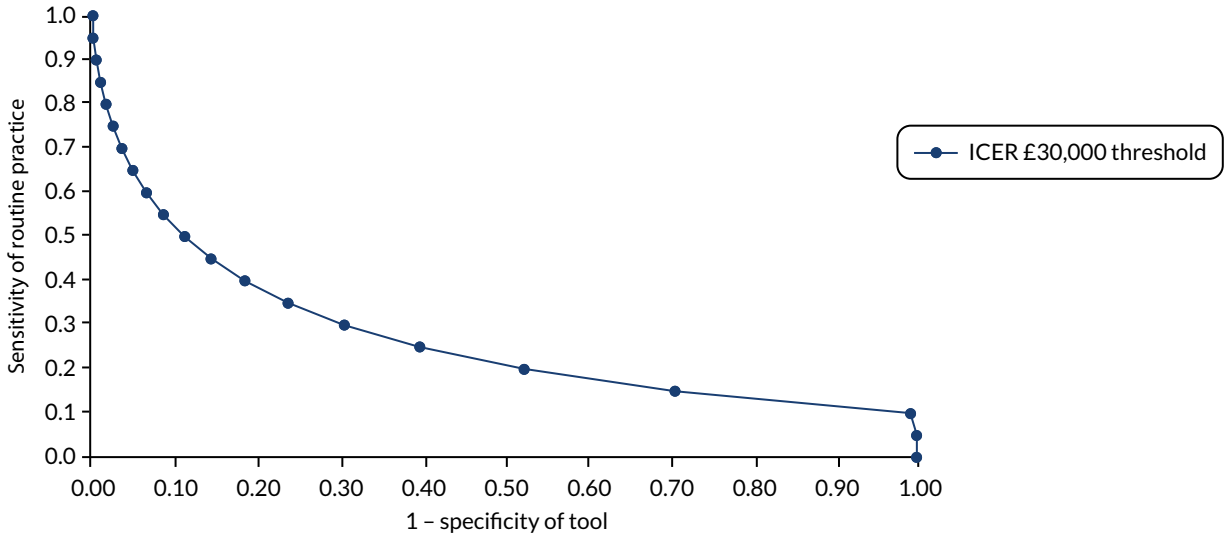


FIGURE 14 Threshold analysis.

Scenario analyses

A key area of uncertainty in the economic analysis of the diagnostic tools is the diagnostic sensitivity of current standard practice. The only evidence available on sensitivity is reported from a retrospective survey,[173] in which a sensitivity of primary care referral in routine practice of 76% is implicit. In the absence of data from this source,[173] we assume the specificity of current practice equals the specificity that corresponds to its sensitivity value in the receiver operating characteristic curve reported for the NICE version 3 referral guidelines.[22] Using these sensitivity and specificity parameters, the intervention produces negligible and possibly negative (discounted) QALY effects (*Table 25*). A similar outcome would arise in younger suspected cancer patients, presenting at 55 years of age. Limiting of the analytical time horizon to 5 years, instead of the 30-year time horizon in our base-case analysis, increases the cost per QALY gained by 73–89%, depending on the tool. Imposing a maximum of 12 primary care visits before referral would have no significant effect on the cost per QALY gained by the tools.

In addition, we explored the uncertainty in the cost-effectiveness of the tools from allowing for a 'send home/wait' comparator strategy, whereby, at the initial visit, the GP sends the patient home without conducting a FIT; from the second visit onwards, they manage the patient as in the 'FIT given to all'

TABLE 25 Scenario analyses

Scenario	RAT			QCancer		
	Incremental cost (£)	Incremental QALYs	Cost per QALY gained (£)	Incremental cost (£)	Incremental QALYs	Cost per QALY gained (£)
Base case	104.54	0.0015	71,863	41.75	0.0015	28,704
Higher prevalence rate (2%)	104.52	0.0020	52,470	42.02	0.0020	21,439
Compliance of 100% with test result	129.87	0.0008	162,183	51.45	0.0009	58,313
Sensitivity of 76% and specificity of 52% for standard practice[a]	52.99	0.0003	167,972	20.85	0.0003	62,368
FIT of 10 µg Hb/g faeces; sensitivity of 92% and specificity of 86%	91.34	-6×10^{-5}	Dominated	35.66	1.6×10^{-5}	2.26M
Age 55 years at presentation	110.08	0.0019	57,767	45.38	0.0019	23,516
No mortality benefit of treatment in stages A/B (HR = 0)	104.45	0.0014	76,930	41.66	0.0014	30,552
Limit CRC-related death risk to 5 years	104.54	0.0018	57,485	41.75	0.0018	23,234
Limit time horizon of analysis to 5 years	104.54	0.0007	148,294	41.75	0.0007	59,902
Restrict maximum number of GP visits to 12	104.54	0.0015	71,979	41.75	0.0015	28,749
Restrict maximum number of GP visits to 12	104.36	0.0013	81,505	41.57	0.0013	32,408
Same CRC stage utilities across disease stages	104.54	0.0017	63,046	41.75	0.0016	25,333
Disutility of colonoscopy (0.0075)	104.54	0.0014	77,211	41.75	0.0014	29,507

HR, hazard ratio.
Note
a Lyratzopoulos *et al.*[173]

strategy. This is a highly uncertain comparator to model as a result of the absence of any information on the rate of return of both CRC and non-CRC patients after an initial visit resulting in neither investigation in primary care nor referral to secondary care. To simplify the analysis, we have assumed that all CRC patients return the following month to seek care for their unresolved symptoms, whereas only a proportion of non-CRC patients do so. Thus, we show how the cost-effectiveness of the tool triage strategies vary with the proportion of CRC-negative patients who return to the GP after the first month. When 0%, 50% and 100% of these patients return, the incremental discounted cost per QALY gained with the RAT relative to this 'send home and wait for second visit before FIT' comparator is £50,605, £37,245 and £23,993, respectively. The corresponding figures for QCancer are £30,472, £17,194 and £4023.

We also explored a modified intervention strategy whereby patients who are above the risk threshold of the tool are referred directly to secondary care, as in the original intervention strategy (see *Figure 2*), whereas those below the threshold are sent home, rather than being subjected to primary care investigations, as assumed in our base-case analysis. In this scenario, the incremental cost per QALY gained with the tools over the FIT-alone strategy was predicted to be £40,140 for QCancer and £96,285 for the RAT.

We also investigated whether or not there would be any possible value of quality-of-life gains from more expeditious referral and diagnosis that would be sufficient to make the tools cost-effective, in a scenario in which more expeditious referrals do not result in survival gains. We found that, even at zero utility values of time spent in the false-negative health state, that is with undiagnosed CRC, the tools would not be cost-effective. At such extremely low values, the ICER for QCancer was £68,102 and that for the RAT was £190,294.

Discussion

In this chapter, we present exploratory analyses based on a simple model of the diagnostic pathway that has as its starting point symptomatic patients presenting to primary care and undergoing an initial clinical assessment, which is the same starting point considered by NICE Guideline Number 12[6] and the Diagnostic Guideline Number 30.[140] We then combine the model with an adapted disease history model for CRC from published studies (see *Chapter 6*), using a shorter cycle length and accounting for the excess risk of mortality due to undiagnosed CRC. By developing these features in our model, we were able to produce predictions in terms of meaningful clinical outcomes. The model was then used to identify the parameters that contribute most to the overall decision uncertainty about the cost-effectiveness of decision tools, which are good candidates for assessment as research priorities using expected value of partial perfect information.[166–168] Our model is expected to undergo further refinement as new evidence emerges, thus becoming an effective live research resource for evaluation and decision-making analysis.

Our analysis of key areas of uncertainty in the economic evaluation of the diagnostic tools led us to the following findings. First, a key determinant for the relative clinical effectiveness and cost-effectiveness of a diagnostic tool-based pathway is the diagnostic accuracy of current standard practice, which is currently unknown. We can only infer what this accuracy might be from the limited data available in a single report of findings from a retrospective survey.[173] Alternatively, as we have also done in this chapter, one may assume that patients are being managed by the results of the FIT, using the diagnostic accuracy estimates for this test available in the literature.[175] This may be a more consistent approach, but limits the scope allowed for accounting for clinician's own judgement in the referral decision-making process.

Second, there is a lack of evidence for evaluating the impact of diagnostic tools in clinical practice. It is not known how these tools would be used in actual practice and how any impact they might have on referral intervals would affect clinical outcomes. We have opted for modelling the expected accuracy of using the diagnostic tools to triage patients for primary care investigation with a FIT, under the strong assumption that the diagnostic accuracy of FITs does not depend on the outcome of the diagnostic tool used for triage.

Third, there is little evidence on the prevalence of cancer in the low-risk population in which the diagnostic tools are intended for use. The value of this parameter is crucial for the cost-effectiveness of using the tool to detect more cancer patients early, which depends on limiting the cost imposed on the overwhelming majority of patients who present with symptoms of suspected cancer but turn out later not to have the disease. Previous analyses, including those informing the NICE Guideline Number 12[6] and Diagnostic Guideline Number 30,[140] have assumed that the CRC prevalence in the low-risk population of interest here is 1.5% based on expert opinion, which, naturally, is highly uncertain. We have thus adopted this value in our base-case analyses. In any case, our exploratory analysis suggested that the specificity of the tools will also determine whether or not these may be used efficiently, without incurring excessive costs of colonoscopies performed in people without cancer, and are sustainable for a universal public health system such as the English NHS.

Fourth, the absence of data on the rate of CRC disease progression prior to diagnosis is also of importance. In our analysis, the slower the rate of progression, the lower the relative clinical effectiveness

and cost-effectiveness of diagnostic tool triage to primary and secondary care investigations, as the tool triage would detect fewer extra cases before they had reached advanced disease stages, for which treatment is less effective, relative to those currently being detected in routine practice. However, the tornado analysis revealed that the sensitivity of current primary care practice, the specificity of the tool and the way clinicians would use the tools are more important sources of uncertainty than the disease progression in the evaluation of these tools.

Fifth, uncertainty regarding the decision-making process currently used in primary care for referring patients to secondary care owing to suspected CRC prevents drawing conclusions about the optimal intervention strategy. When we compared the intervention with a strategy of sending symptomatic patients home on the first visit and using the FIT-alone strategy for those patients who return with unresolved symptoms for a second visit, the proportion of non-CRC patients who return for the second visit determines whether or not a diagnostic tool approaches levels of cost-effectiveness. Thus, the economic value of the tools will depend on the clinical effectiveness of delaying access to colonoscopy as the alternative strategy for identifying patients who need referral, which may avoid unnecessary use of health-care resources and health risks of invasive investigations, at the cost of putting some patients with CRC at risk by delaying their identification and treatment.

Another limitation in our analysis is the lack of evidence on diagnostic accuracy from any direct comparative study of the modelled strategies. Thus, our sensitivity and specificity values are derived from single studies that are likely to be different in terms of population, as reflected, for example, in the underlying prevalence rate, or study design; for example, the evidence from one of the tools is derived from a case–control study, whereas the evidence for the other tool was obtained from a cohort study. Therefore, we cannot directly compare QCancer and RATs, as this is likely to lead to biased results.

In addition, we have not modelled the heterogeneity in current standard practice; therefore, we have not accounted for the variation in the costs and benefits of using the tools depending on local situations. It is conceivable that practices with robust referral systems may have little to benefit from using the tools, whereas areas without a sound referral system may find them beneficial. The model assumes 100% compliance of GPs with the strategies (both FITs and decision tools). However, it is clear that, currently, not all GPs use decision tools. More research is therefore needed to understand how compliance with strategies varies across localities.

The model does not account for the resource constraint effects imposed by limited capacity to do colonoscopies. Thus, although our model accounts for the benefits of faster access to colonoscopy, it does not include the opportunity costs of those patients whose access to colonoscopy will be delayed as a consequence of further demands placed on such service with limited capacity. This is an area of further research.

Conclusion

The decision-analytic model was developed to analyse the uncertainty inherent in the current evidence base, and to ask questions about the probable impact of the diagnostic tools, given the current lack of evidence. The model showed the importance of accounting for the excess mortality effects of CRC in symptomatic patients who are undiagnosed, and extends previous models in CRC by explicitly modelling a simple diagnostic referral pathway in primary care. Despite the high degree of uncertainty on evaluating the effect of the diagnostic tool on primary care referral of suspected CRC and health outcomes, our model has helped us to identify the parameters for which uncertainties contribute most to the overall decision uncertainty about the cost-effectiveness of decision tools, namely the sensitivity of current standard practice, the specificity of the diagnostic tool, the prevalence of CRC in the relevant primary care patient population and the cost of colonoscopy.

In the concluding section, we return to the original questions about the probable impact of the diagnostic tools on the costs and effects for patients presenting with symptoms of suspected CRC. For each question we summarise the initial findings from the model and highlight where the uncertainties still remain.

- What are the possible impacts on patient quality of life or survival if use of diagnostic tools reduces time to diagnosis?
  The model-based analysis suggests that the diagnostic tools would reduce the number of CRC patients who die without a diagnosis from 12% to 8%. The increased accuracy of the tool relative to the comparator also reduces the referral interval, which equals a reduction in the time to diagnosis and treatment by assumption. Owing to the challenges of measuring the effect on health-related quality of life post diagnosis, the amount of this type of benefit is highly uncertain. Overall, our modelled analysis suggests that, depending on the sensitivity of current practice, the diagnostic tools may extend life expectancy by somewhere between 4 days and 2 months among symptomatic patients aged 70 years presenting to their GP with undiagnosed CRC, relative to a hypothetical current practice of management according to results of a FIT alone. This range of variation in benefit highlights the importance of establishing current practice and its primary care diagnostic accuracy.
- Will the benefit to cancer patients identified earlier by diagnostic tools outweigh any disutility in extra patients referred for further investigation who do not have cancer?
  In the absence of evidence, our analysis suggests that the tools may help to improve diagnostic accuracy in primary care, and result in earlier referrals to secondary care. If the reduction in the referral interval results in a shortening of the diagnostic interval of the same magnitude, our model predicts that the amount of benefit to patients with CRC identified earlier outweighs the loss in quantity and quality of life in those patients referred for further investigation who do not have CRC. However, although the benefit to cancer patients would more than compensate the risks to life from exposing the overwhelming majority of patients without cancer to the risks of colonoscopy, it is unlikely to justify the additional health-care costs associated with such exposure.
  A sensitivity analysis was used to explore the uncertainty around the estimated cost-effectiveness results and to highlight the parameter values that were driving the results. The sensitivity analysis revealed that the cost-effectiveness results were particularly sensitive to uncertainty around the sensitivity of current standard practice and the specificity of the tool. In a further threshold analysis, we explored the levels of sensitivity of current standard practice and the specificity of the tool in which the decision tool would meet the £30,000 cost-effectiveness threshold. At the levels of sensitivity of $\geq$ 50% that have been reported for a FIT of 20 µg Hb/g faeces,[175] the minimum level of specificity needed by the tool to serve as a cost-effective triage intervention in primary care is 0.89. The limited available data on current practice in the UK suggests that the sensitivity of primary care referrals is already too high for the diagnostic tools to make a significant impact on CRC patient survival. Although our base-case analysis produced an estimate of 2 months of life extension with the tools when compared with a FIT of 20 µg Hb/g faeces, this gain is reduced to 2 weeks when the sensitivity value derived from the number of visits before CRC diagnosis in routine practice is used, and to 4 days against a FIT of 10 µg Hb/g faeces. Given the very low prevalence of CRC in the relevant primary care population, this amount of benefit may be insufficient to outweigh the risks to life from exposing a vast number of patients without CRC to colonoscopy. Furthermore, whether or not the tools are cost-effective will depend on the prevalence and the sensitivity of current referral practice and the specificity of the tool.
- How big an improvement in quality of life would be needed to warrant the use of these tools if there are no survival impacts associated with the diagnostic tools?
  We investigated this question by assuming that the increased accuracy of the tool relative to the comparator does not result in any survival benefit but would avoid any loss in quality of life arising from false-negative results, whereas the quality of life of a false-positive result remains unaffected. The benefits gained from the tools are then largest when the losses from false-negative results are the biggest, as would occur when quality of life of these false-negative patients falls to zero.

Our analysis, however, shows that even preventing these extreme losses in quality of life for false-negative patients is insufficient to make the diagnostic tools cost-effective (the ICER for QCancer is then £68,102 and for the RAT is £190,294). Thus, without survival benefits, the quality-of-life benefits of the tools alone would be unlikely to be sufficient to justify their costs.

- Could a cancer diagnostic tool be considered cost-effective if it reduces the period of extreme anxiety by expediting investigation and management in patients, even if it made no impact on patient outcome?
  The model results do not account for any negative impact of referrals on a patient's quality of life due to anxiety, or the impact of colonoscopy on patient quality of life, for which we found no evidence in the literature. Therefore, we cannot provide an answer to our study question of whether or not a cancer diagnostic tool reduces the period of extreme anxiety associated with an uncertain CRC diagnosis by expediting investigations and management. Answering this question would require prospective studies measuring the outcomes of patients in primary care, but it is questionable whether or not such a study is feasible given the difficulty to define the patient population and the low number of cases expected in a single practice each year. Nevertheless, our exploratory analyses suggest that, even if plausible reductions in the diagnostic interval brought about by the use of the tools were to equate to reductions in the amount of time patients experience extreme anxiety, the tool would not be cost-effective without extending patient survival.
- Where are the gaps in the evidence base and where is more research needed?
  The updated review showed that the statistical relationship between diagnostic delay and cancer patient outcomes is not suitable for deriving valid estimates of the effect of expeditious CRC diagnosis on cancer stage or patient survival. In particular, the documented non-monotonic (e.g. U-shaped) relationship between diagnostic interval and mortality does not imply an effect mediated through the impact of diagnostic interval on stage of diagnosis. Given the lack of valid surrogate or clinical outcome data for building an economic model of diagnostic tools or decision-aid models in primary care, our cost-effectiveness analysis was limited to an exploration of the expected effects of time delays to diagnosis associated with the relative accuracy of the diagnostic tools. This limited aim will naturally be subject to more sources of uncertainty than an analysis based on clinical surrogate outcomes, but reflects a pragmatic approach given the quality of the data available.
  In the absence of direct clinical effectiveness data, we assumed a relationship between the sensitivity of both the diagnostic tool and current routine referral practice and the time to diagnosis. In the model, the time from symptom presentation to diagnosis consists of (1) the time from presentation in primary care to referral for investigation (the referral interval; see *Chapter 5*) and (2) the time from referral to specialist care to diagnosis. We then assumed that the diagnostic algorithm affects only the referral interval, while the time from referral to diagnosis and treatment remains fixed. Thus, in our analysis, we assume that the improvement of primary care diagnostic accuracy with the tool results in fewer visits to the GP before referral, and that the reduction in the accompanying time to referral is equal to the reduction in the time to diagnosis and treatment. We have explicitly modelled the average time to referral and diagnosis and treatment, based on the number of repeat appointments to make the mechanisms of effect transparent and open to scrutiny. The relationship between the number of repeat appointments and the time to referral is supported by evidence from a retrospective patient experience survey.[173] More research is needed on the impact of the different strategies on the number of GP appointments before referral and diagnosis.
  The review highlighted that there was a lack of direct comparative accuracy evidence between the strategies/tests involved in the model. Our model is therefore based on indirect comparisons of single-arm accuracy studies and assumptions about how the tools would be used to triage patients for investigations in secondary care. There is limited evidence on the diagnostic accuracy of decision aid tools in the UK low-risk population of interest, and the methods used to evaluate the tools that have any relevant evidence, QCancer and the RAT, differ in terms of how the relevant low-risk population is defined, that is 'top 1 to 9% risk' versus 'low risk but not no-risk',[75] and the study design used to develop or evaluate them, that is observational cohort versus case–control, respectively.
  As QCancer was developed from a large cohort from electronic medical records, data were

not prospectively collected and may thus lack relevant predictors to primary care practitioners. The RAT, in contrast, warrants validation in a large representative UK primary care population, because evidence on its diagnostic accuracy is based on a retrospective patient selection design. In addition, the added value of these tools is highly uncertain as the status quo has not been defined or rigorously evaluated, as reflected by the poor quality of the evidence informing the latest NICE guidance (Diagnostic Guideline Number 30[3]), including the FIT, which has been recommended by such guidance but is not yet universal at the time of writing.

More research is also needed to verify the relationship between diagnostic delay and cancer patient outcomes. This is a difficult area to study because of unobserved confounding factors, but one for which access to electronic medical health records with information to construct detailed symptoms libraries has opened new opportunities for fruitful research.

The sensitivity analysis revealed that the cost-effectiveness results were particularly sensitive to uncertainty around the diagnostic accuracy of current standard practice and the specificity of the tool. We assume that all patients in the population of interest, who we defined as those with 'low risk but not no-risk symptoms',[75] are akin to the CAPER study population.[75] This patient population includes the majority of symptomatic patients who do not present with high-risk symptoms such as rectal bleeding or severe anaemia, and therefore are not diagnosed through the 2WW referral pathway. The best available evidence on accuracy of the tool is for the QCancer tool, as it is based on a large cohort analysis. However, even these sensitivity and specificity parameters were not for the same population of interest here, as QCancer values include patients with red-flag symptoms, who may be deemed appropriate for 2WW referrals. Therefore, in our model, we calculated the sensitivity and specificity of QCancer for the top 10% risk threshold after excluding the top 1% risk observations from the published data[77] to be 61% and 91%, respectively. In a further threshold analysis, we explored the levels of accuracy of current standard practice and the specificity of the tool in which the decision tool would meet the £30,000 cost-effectiveness threshold. At the levels of sensitivity of $\geq$ 50% that have been reported for a FIT of 20 µg Hb/g faeces,[175] the minimum level of specificity needed by the tool to serve as a cost-effective triage intervention in primary care is 0.89. More research is clearly needed on the specificity of these tools among the population of interest. In addition, we have not modelled the heterogeneity in current standard practice; therefore, we have not accounted for the variation in the costs and benefits of using the tools depending on local situations. It is conceivable that practices with robust referral systems may have little to gain from using the tools, whereas areas without a sound referral system may find them beneficial. The model assumes 100% compliance of GPs with the strategies (both FITs and decision tools). However, it is clear that, currently, not all GPs use decision tools. More research is therefore needed to understand how compliance with strategies varies across localities.

The model does not account for the resource constraint effects imposed by limited capacity for colonoscopies. Thus, although our model accounts for the benefits of faster access to colonoscopy, it does not include the opportunity costs of those patients whose access to colonoscopy will be delayed as a consequence of further demands placed on such service with limited capacity. This is an area of further research.

Finally, there were insufficient data to evaluate one of the a priori aims of the model, which was to explore whether or not the tools will become cost-effective if they reduce the period of extreme anxiety, even if they made no further impact on patient outcome. Answering this question would require prospective studies measuring the outcomes of patients in primary care, but is questionable whether or not such a study is feasible given the difficulty to define the patient population and the low number of cases expected in a single practice each year.

Chapter 8 General practitioner survey

Introduction

Diagnosing cancer quickly once patients become symptomatic remains an important priority in the UK.[185–187] A key part of the UK's strategy for early cancer diagnosis was to lower the threshold for fast-track referrals of symptomatic patients for investigation from an estimated 5% to an explicit 3% risk of undiagnosed cancer. This was implemented via the 2015 revision of NICE guidelines on recognition and referral for suspected cancer.[6] A potential source of diagnostic delay remains the UK's 'gatekeeper' system, whereby a GP decides whether or not a symptomatic patient meets the criteria for investigation for suspected cancer.[188] Indeed, fast-track referrals for cancer are less likely when patients present with 'low-risk but not no-risk'[75] symptoms of cancer than when they present with 'alarm' symptoms.[189]

A number of models and algorithms have been developed to quantify the risk of an undiagnosed malignancy in symptomatic patients (see *Chapters 3* and *4*).[42,44] Some of these models have been translated into cancer decision support tools for clinical use, helping GPs to identify patients who warrant investigation for suspected cancer.

Systematic review 1 (see *Chapter 3*) identified two tools that are available to support doctors assess the risk of undiagnosed skin cancer.[36–38] These tools are adapted to Australian guidelines and have limited applicability in the UK. For the internal cancers, there are two main types of cancer decision support tools available in the UK: RATs and QCancer. RATs are available for a number of specific cancer sites, and use symptoms and test results to estimate the risk that a patient has an underlying cancer.[51,66,69–71,73,81,83,90] They were identified in SR1 (see *Chapter 3*) as having been evaluated in their mouse mat and desktop forms. QCancer tools estimate the risk of undiagnosed cancer based on symptoms, test results and patient risk factors. They are available for specific cancer sites,[32,52,80,82,84,85] and also in sex-specific, person-centred forms, which estimate an individual's risk that they have a number of different cancers.[53,54] QCancer tools were identified as validated cancer prediction models in SR2 (see *Chapter 4*).

In 2013, Macmillan Cancer Support, BMJ Informatica, the Department of Health and Social Care, and Cancer Research UK collaborated to incorporate RATs and QCancer into GP IT systems, renaming them collectively as 'electronic cancer decision support tools'. RATs are fully integrated with the GP IT software INPS Vision (In Practice Systems Ltd, London, UK) and will soon be available via SystmOne (The Phoenix Partnership, Leeds, UK). QCancer is available via EMIS Web (EMIS Health, Leeds, UK), and QCancer tools are also freely available on the internet (www.qcancer.org). NHS Digital data indicate that, together, EMIS Web and Vision had 62% of the market share of GP IT systems in 2015.[190] In addition, RATs are available in mouse mat and flip chart forms; in January 2012, they were distributed to all 10,000 general practices in England as part of the National Cancer Action Team's Supporting Primary Care Project.[185] Information about QCancer and RATs is available via the Cancer Research UK and Macmillan Cancer Support websites.[191,192]

The electronic cancer decision support tools have three main functions:

1. alert/prompt – cancer risk scores appear automatically on opening a patient's record
2. cancer risk assessment – GPs can request a patient's cancer risk score, using the symptom checker
3. searches/report – GPs can routinely search records and produce summaries indicating patients who may need follow-up ('safety-netting').

Existing research

There is little existing research on the clinical effectiveness and cost-effectiveness of cancer decision support tools (see *Chapters 3* and *4*).

Independent and external validations of the QCancer algorithms for colorectal,[15] pancreatic,[14] renal tract,[91] gastro-oesophageal[88] and ovarian[89] cancers concluded that these tools are useful for identifying undetected cases of these cancers in primary care (see *Chapter 4*). The performance of the CRC RAT was compared with that of the 2005 NICE guidelines for referral for suspected cancer.[172] The RAT (area under the receiver operating characteristic curve 0.91, 95% CI 0.89 to 0.93) performed better than the 2005 NICE guidelines (area under the receiver operating characteristic curve 0.75, 95% CI 0.72 to 0.79) in discriminating patients with undiagnosed CRC.[22,172] No comparison with the current NICE guidelines[6] has been carried out; however, these guidelines draw heavily on an evidence base dominated by RATs and QCancer publications.

A cohort study compared the numbers of cancer investigations and diagnoses before and after the introduction of CRC and lung cancer RATs in primary care in the UK. The introduction of RATs was associated with increased diagnostic activity and additional diagnoses of lung and colorectal cancer.[13] By contrast, a 2 × 2 design trial of a GP intervention, which included the colorectal and lung cancer RATs, did not report any evidence that the GP intervention was associated with faster time to diagnosis of cancer in rural Australia.[35] The clinical utility of cancer decision support tools in primary care remains uncertain.[42] No studies have investigated the association between the use of cancer decision support tools and the use of the urgent referral pathway for suspected cancer.

Qualitative studies of cancer decision support tools in UK primary care suggest that they are more likely to be embedded in clinical practice if they are perceived to support, but not supersede, a doctor's clinical judgement, and with better training and improvements in design. In particular, their use of screen prompts added to the barrage of alerts generated by GP software systems, thereby increasing the risk of 'prompt fatigue' and disengagement. Positive aspects of the tools included raising awareness of cancer and of its symptoms, and prompting GPs to reflect on their decision-making around referrals.[13,16,17] A recent qualitative study aimed to improve the understanding of how cancer decision support tools are used by GPs. It analysed responses from a convenience sample of 126 GPs, most of whom worked in the South West Peninsula. The study reported that 18.3% of GPs used either a RAT or a QCancer, and that awareness of these tools was low (Chisnell *et al.*, submitted).

No studies have quantified the proportion of general practices and GPs in the UK that have access to and use cancer decision support tools. This is essential for increasing our understanding of the tools' clinical effectiveness and cost-effectiveness.

Research aims

The primary aims of this study were to quantify the proportion of:

- general practices and GPs in the UK that have access to the two main types of cancer decision support tools, either through a paper-based desktop tool or via their IT software (i.e. RATs and QCancer)
- general practices where at least one GP reports using these tools.

Secondary aims of the study were to investigate the association between access to cancer decision support tools and cancer diagnostic activity. Diagnostic activity was assessed using two indicators. The first is a diagnostic process indicator, namely the age- and sex-adjusted number of 2WW referrals per 100,000 head of population for the general practice. The second is a diagnostic outcome indicator, namely the proportion of patients referred via the urgent 2WW pathway who are subsequently diagnosed with cancer.[193]

Methods

A cross-sectional postal survey was designed and conducted in line with conduct and reporting guidelines for surveys.[194]

Questionnaire design

The questionnaire was drawn up by the lead researcher (SP) and was refined following feedback from Professors Hamilton, Hippisley-Cox and Coupland (the originators of RATs and QCancer), and members of the research team. Images of the paper-based tools and screenshots of the electronic cancer decision support tools were included in the questionnaire (courtesy of In Practice Systems Ltd and EMIS Health) to ease their identification by the participants. Copies of the questionnaire, covering letter and information sheet are included in *Appendix 6*. Each questionnaire included the general practice identifier, but not the name of the responding GP.

The questionnaire was finalised following feedback from five local GPs in terms of its utility, clarity and design. The final questionnaire comprised 10 questions. Questions 1–3 asked GPs about desktop RATs, including whether or not they were available and how likely they were to be used. Questions 4–8 asked about electronic cancer decision support tools (i.e. RATs and QCancer), again focusing on availability and likely use, and training. Questions 9 and 10 asked how many years the GP had been practising and how many sessions per week they worked at the practice. Participants were also asked to select, from a list, aspects of the tools that they found helpful. The contents of the list were informed by the qualitative study findings on the positive aspects of the lung cancer and CRC tools discussed above.[13,16,17]

Sample

The survey was conducted at the practice level, reflecting how decisions are made about the IT software and whether or not to download or activate electronic cancer decision support tools.

The target population was general practices in the UK and the GP partners/principals, sessional GPs (including salaried and locum GPs) and GP registrars working there. GPs who had retired or who were not currently practising were excluded.

We estimated that a sample size of 392 general practices would be large enough for a 95% CI to have a margin of error of no more than 5%. We assumed a response rate of 40%, which is considerably less than the mean value of 61% (95% CI 59% to 63%) reported in a pooled analysis of 361 postal surveys published between 2000 and 2009.[195]

Survey administration

The postal survey was administered by Binley's (Basildon, UK; www.binleys.com). This commercial company was selected because it maintains a database of all 46,000 GPs in the UK, and all general practices, which is verified every 6 months. This enabled us to obtain a random probability sample of 975 general practices. Questionnaires were sent to all the GPs ($n = 4350$) and GP registrars ($n = 250$) working in these practices. Copies for GP registrars were colour coded, to facilitate separate analyses of their responses, as these GPs are not on the General Medical Council's GP register. The questionnaires were sent on 5 July 2017, with a follow-up questionnaire pack sent to non-responding practices on 2 August 2017. The data collection stopped on 15 November 2017. To incentivise participation, a charitable donation of £7.50 was made per completed questionnaire, capped at the first 400. The £3000 donation was shared equally between Cancer Research UK and Macmillan Cancer Support.

Data entry and coding

The data were entered into a Microsoft Excel® (Microsoft Corporation, Redmond, WA, USA) spreadsheet by Binley's, and double-checked for accuracy by a second data entry analyst at Binley's. The questionnaire did not include a free-text comments section; however, any comments that were

written on the returned questionnaires, or sent by e-mail or telephone, were recorded and are reported in *Appendix 6*. The paper copies of the questionnaire were sent to the University of Exeter for reference and archiving. The numbers of general practices and individual GPs in a practice who failed to respond were recorded. If one GP reported that they had access to cancer decision support tools, it was assumed that this was true for all other GPs at that practice.

Analyses

Descriptive statistics are reported for the responses to each question, along with the numbers of missing data. GP-level data were used to estimate the proportion (95% CI) of GPs in the UK with access to the tool. Responding practices were categorised in two ways: first, by whether or not at least one GP had access to a tool, and, second, by whether or not at least one GP used a tool. From this, we estimated the proportion (with 95% CI) of UK general practices with access to a cancer decision support tool, and the proportion of UK general practices that use the tool.

Two subanalyses examined data from general practices in England only.

Ordinary least squares regression analyses tested the association between access to cancer decision support tools and a practice-level diagnostic process indicator, namely the number of urgent 2WW referrals for suspected cancer. The dependent variable was the number of sex- and age-adjusted 2WW referrals for suspected cancer per 100,000 head of population. The independent variable was a dummy variable indicating whether or not at least one GP at the practice had access to a tool. The analysis adjusted for the general practice's Index of Multiple Deprivation (IMD). This is a population-weighted composite measure of 37 different measures of deprivation, taking a value between 0 and 100. The higher the score, the higher the level of deprivation.[196,197]

Ordinary least squares regression was also used to test the association between access to cancer decision support tools and a practice-level diagnostic outcome indicator, namely the proportion of 2WW referrals that result in a cancer diagnosis (also known as the conversion rate). This also adjusted for the practice-level IMD.[196,197]

The 2WW and conversion rate data and practice-level IMD data for the period 2016–17 are published by Public Health England and are publicly available at http://fingertips.phe.org.uk/profile/cancerservices (accessed December 2019). Survey responses were mapped to Public Health England data using the general practice identifier.

Results

Sample characteristics

Responses were received from 473 GPs and from three GP registrars based in 227 practices. Responses from the GP registrars were not analysed separately, because of the small number of GP registrars providing responses. The response rate at the practice level was 23.3%; at the practitioner level, it was 10.3%. The responding practices had a median of 6 [interquartile range (IQR) 4–8] GPs, of whom a median of 2 (IQR 1–3) responded to the survey. The mean within-practice response rate was 43.7% (95% CI 39.3% to 48.1%). Unprompted comments written on incomplete surveys returned by practice managers, or messages received by e-mail or telephone (see *Data entry and coding*) indicated that lack of time ($n = 12$) and lack of awareness of the tools ($n = 6$) were the most common reasons for non-response.

Of the 476 respondents, 294 (61.8%) had been practising as a GP for ≥ 11 years, and 299 (62.8%) worked between five and eight sessions per week (*Table 26*).

TABLE 26 Numbers of years in practice and number of consultation sessions worked per week

Demographic information	GPs	
	n	%
Years in practice		
< 1	20	4.2
1–5	57	12.0
6–10	57	12.0
11–20	145	30.5
21–30	111	23.3
> 30	38	8.0
Missing	48	10.1
Total	476	100.00
Number of consultations sessions worked per week		
1–2	10	2.1
3–4	70	14.7
5–6	140	29.4
7–8	159	33.4
9–10	40	8.4
> 10	5	1.1
Missing	52	11.0
Total	476	100.0

EMIS Web was the most frequently used IT software (96/227, 42.3%), followed by TPP SystmOne (74/227, 32.6%) and then INPS Vision (32/227, 14.1%). These relative frequencies largely reflect the national market share (*Table 27*).

The IMD profile of the sample is generally representative of that of the national population (*Figure 15*).

TABLE 27 General practice software

General practice IT software (eCDS tool)	Responding general practice		National market share (%)
	n	%	
EMIS Web (QCancer)	96	42.3	52.4
TPP SystmOne (none)	74	32.6	33.0
INPS Vision (RAT)	32	14.1	9.9
Egton Medical Information Systems PCS (none) (EMIS Health)	11	4.9	0.01
Egton Medical Information Systems LV (none) (EMIS Health)	5	2.2	0.02
Other (none)	6	2.2	3.3
Microtest (none)[a]	3	1.3	1.4
Total	227	100	100

a Microtest Health, Bodmin, UK.

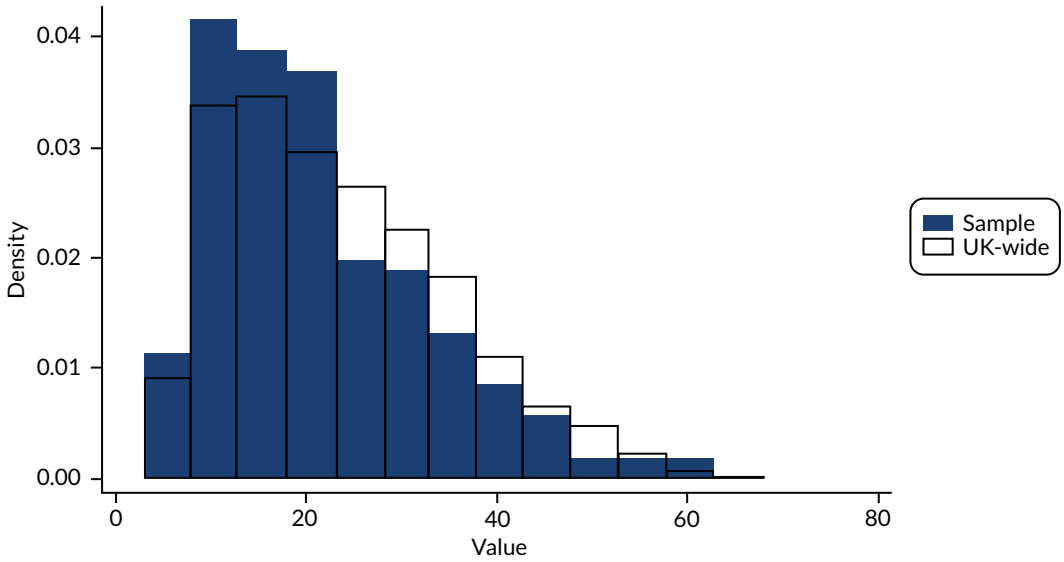


FIGURE 15 Index of Multiple Deprivation profiles.

Paper-based risk assessment tools

Access to a paper-based RAT in mouse mat or flip chart form was reported by 63 of the 476 (13.2%) GPs. At the practice level, at least one GP reported access to a tool in 51 of the 227 (22.5%, 95% CI 17.2% to 28.5%) responding general practices (*Table 28*). The 'other' tools are listed in *Appendix 6*; all were national guidelines, or summaries thereof, rather than cancer decision support tools.

Of the 63 GPs with access to a mouse mat or flip chart, 39 (61.9%) reported that they were unlikely or very unlikely to use it, and 19 (30.2%) reported that they were likely or very likely to use it (see *Table 28*).

TABLE 28 Responses to questions asking about availability of risk assessment tools, and likely use of paper-based tools (mouse mat or flip chart)

	GPs	
Question category	***n***	**%**
Format of RAT		
Mouse mat or flip chart	63	13.2
Other	30	6.3
None of these	326	68.5
Missing, not answered	57	12.0
Total	476	100.0
Likelihood of using paper-based RAT		
Very likely	5	7.9
Likely	14	22.2
Unlikely	29	46.0
Very unlikely	10	15.9
Missing, not answered	5	7.9
Total	63	100.0

The most popular probable uses of the paper-based RATs were for:

- assessing cancer risk in patients with non-specific ($n = 28/63$, 44.0%) or multiple ($n = 25/63$, 40.0%) symptoms
- increasing the GP's certainty of decision-making ($n = 25/63$, 40.0%).

Other likely uses were less popular, namely:

- discussing cancer risk ($n = 19/63$, 30.2%)
- reassuring anxious patients ($n = 16/63$, 25.4%)
- increasing awareness of cancer as a possible diagnosis ($n = 17/63$, 27.0%) and awareness of cancer symptoms ($n = 10/63$, 15.9%)
- prompting referrals the GP would not otherwise have made ($n = 13/63$, 20.6%), or investigation ($n = 9/63$, 14.3%)

The option 'none of these reasons' was selected by 8 of the 63 (12.7%) GPs.

Electronic clinical decision support tools

The electronic clinical decision support tool was downloaded or activated on the IT system of 58 of the 476 responding GPs (12.2%) (*Table 29*), which equates to a practice level of 42 out of 227 (19.0%, 95% CI 14.0% to 24.6%). Practices using EMIS Web and INPS Vision were equally likely to have downloaded/activated the software (EMIS Web: $n = 32/96$, 33.3%; INPS Vision: $n = 10/32$, 31.3%) (*Table 30*). Of the 476 GPs, 174 (36.6%) were unaware of the electronic tools, and 39 (8.2%) reported that they would like to have them but they are not available for their system.

Of the 58 GPs with access to the electronic clinical decision support tools, 17 (29.3%) reported having integrated it into their practice, and nine (15.5%) reported receiving training. The proportion who had both received training and had integrated the tool into their practice was low (5/58, 8.6%). At the practice level, training had been received by at least one GP in six of the 42 (14.3%) practices with access to the tool. The tool was integrated into the practice of at least one GP in 15 of the 42 (35.7%) practices.

The 'alert prompt' and 'symptom checker' functions were ranked as being the most useful by 16 (27.6%) and 14 (24.1%), respectively, of the 58 GPs with access to the tool. The 'searches report' was ranked least useful by 12 of the 58 GPs (20.7%)

TABLE 29 Responses to question asking about eCDS tool availability

Option	Frequency	
	n	%
Unaware of eCDS	174	36.6
eCDS is downloaded/activated for my IT system	58	12.2
eCDS is available for my IT system, but my practice has not downloaded/activated it	23	4.8
eCDS is available for my IT system, and my practice has plans to download/activate it in future	6	1.3
To my knowledge, eCDS is not available for my IT system	108	22.7
eCDS is not available for my IT system but I would like to have it	39	8.2
Missing, not answered	68	14.3
Total	476	100.0

TABLE 30 General practices with access to the electronic clinical decision support tools

General practice IT software	Access to electronic clinical decision support, *n* (%)		Total (*n*)
	No	Yes	
EMIS Web (QCancer)	64 (66.7)	32 (33.3)	96
EMIS LV (none)	5 (100.0)	0 (0.0)	5
EMIS PCS (none)	11 (100.0)	0 (0.0)	11
INPS Vision (RAT)	22 (68.8)	10 (31.3)	32
TPP SystmOne (none)	73 (98.7)	1 (1.3)[a]	74
Microtest (none)	3 (100.0)	0 (0.0)	3
Other (none)	5 (80.0)	1 (20.0)[a]	6
Total	183 (80.6)	44 (19.4)	227

a Practices that do not use either EMIS Web or INPS Vision cannot have access to electronic clinical decision support tools, so these responses are assumed to be errors on the part of the GP answering the question and are not included in the analysis.

Functions of the electronic clinical decision support tools that were considered helpful included:

- assessing cancer risk in patients with non-specific symptoms (20/58, 34.5%), or with multiple symptoms (21/58, 36.2%)
- discussing cancer risk with a patient (17/58, 29.3%).

Other functions were less popular:

- increasing the awareness of cancer as a possible diagnosis (13/58, 22.4%)
- reassuring anxious patients (10/58, 17.2%)
- prompting referrals that would otherwise have not have made (8/58, 13.8%)
- increasing the certainty of clinical decision-making (7/58, 12.1%)
- increasing the awareness of cancer symptoms (7/58, 12.1%).

Discussing investigation with symptomatic patients was not a popular function (3/58, 5.2%). Approximately one-third of GPs opted to select none of the functions listed (17/58, 29.3%). When asked how likely they were to use the electronic clinical decision support tool to assess a patient whose symptoms may be caused by cancer, 39 (67.2%) reported that they would be unlikely or very unlikely to do so.

Combined tools

Of the 476 GPs, 112 (23.5%, 95% CI 19.7% to 27.6%) had access to a cancer decision support tool in either paper or electronic format, or both. At the practice level, this equates to at least one GP with access in 83 of the 227 practices (36.6%, 95% CI 30.3% to 43.1%). Of the 227 general practices, 38 (16.7%, 95% CI 12.1% to 22.2%) contained at least one GP who had access to the tools and was likely or very likely to use them.

Association between the use of tools and 2-week-wait referral activity

Regression analyses were carried out for the 173 practices in England, for which Public Health England publish age- and sex-adjusted 2WW referral and conversion rates. Of these 173 practices, 68 had access to either a paper RAT or an electronic clinical decision support tool. There was no difference in the mean 2WW referral rate between practices that do and practices that do not have access to either type of tool, after adjusting for IMD (mean difference 1.8 referrals per 100,000 head of population, 95% CI −6.7 to 10.3 per 100,000 head of population) (*Table 31*).

TABLE 31 Regression analysis output: dependent variable sex- and age-adjusted urgent 2WW referrals per 100,000 population[a]

	Referral rate (per 100,000)		
	Mean difference	95% CI	*p*-value
Availability of tool (yes/no)	1.81	−6.72 to 10.34	0.676
IMD	0.63	0.25 to 1.01	0.001

a $n = 173$; $R^2 = 0.06555$.

The normal probability plot of the residuals was approximately linear (data are not shown), supporting the assumption of linear regression that the error terms are normally distributed.

Access to either type of tool was not associated with a change in the proportion of 2WW referrals that resulted in a diagnosis of cancer (the conversion rate), after adjusting for IMD (mean difference −0.2, 95% CI −1.0 to 0.6) (*Table 32*).

Discussion

To the best of our knowledge, this is the first UK-wide survey of the availability of cancer decision support tools. Based on the responses received, cancer decision support tools, in either paper or electronic format, are available to GPs in approximately one-third (36.6%, 95% CI 30.3% to 43.1%) of UK practices. The proportion of general practices where at least one GP had access to the tools and was likely or very likely to use them was 16.7% (95% CI 12.1% to 22.2%). There are no current plans to re-release the paper-based tools, so the expectation is that electronic clinical decision support tools will become the norm. For this reason, it is useful to report separately the availability of electronic clinical decision support tools in UK general practices, which was 19.0% (95% CI 14.0% to 24.6%). Currently, the tools are available only via EMIS Web and INPS Vision, and approximately one-third of the practices using these software systems had opted to download or activate them. The software will shortly be integrated into SystmOne, which is estimated to have 33% of the market share in the UK. Between them, EMIS Web, SystmOne and INPS Vision represent > 95% of the general practice software systems available;[190] therefore, in the near future, it is reasonable to assume that nearly 100% of GPs will have access to the electronic clinical decision support tools, should they choose to download/activate them.

There was no evidence that access to the tools was associated with a change in either the rate of 2WW referrals or the proportion of those who were referred transpiring to have cancer.

Strengths and limitations

The main strength of this study is that we used random probability sampling methods, allowing for every general practice in the UK to have an equal chance of receiving the survey. This increases the

TABLE 32 Regression analysis output: dependent variable sex- and age-adjusted conversion rate (i.e. proportion of 2WW referrals that result in a diagnosis of cancer)[a]

	Proportion of 2WW referrals resulting in a cancer diagnosis		
	Mean difference	95% CI	*p*-value
Availability of tool (yes/no)	−0.18	−0.97 to 0.62	0.663
IMD	−0.05	−0.09 to −0.02	0.005

a $n = 172$; $R^2 = 0.0511$.

generalisability of our results to the UK. Further strengths include the thorough preparation and testing of the survey before distribution to ensure that it was short, clear and user friendly, including images for easy identification of the cancer decision support tools in question. Postal surveys, with personalised letters and prepaid return envelopes, plus a financial incentive by way of a charity donation, were also used, as recognised methods of increasing the response rate.[198]

Despite these steps, our sample size fell short of the intended size, resulting in wider CIs than planned. The main reason for the low response rate was probably large GP workloads, as volunteered by practice managers and reported elsewhere.[199] The potential for responder bias is an important issue to consider here, particularly because the response rate was low. If responders were more likely to have access to, and be engaged with, the tools than non-responders, then our results will be overestimates. One way to explore this is to look at the number of practices with access to computer systems that support the electronic tools to see if they were over-represented in our sample. As seen in the results, the electronic tools were available in only two (Vision and EMIS Web) of the three (Vision, EMIS Web and SystmONE) main general practice IT systems in the UK. In our sample, ≈ 57% of practices had Vision and EMIS systems; this is very similar to the national picture of 62%, and would suggest that the response rate was not related to access to the tools.

Our subanalyses did not provide any evidence of an association between access to the tools and practice-level diagnostic activity. This is perhaps not surprising, given that the electronic tools, in particular, have been available for a only short time and our survey suggests that they are not yet embedded in clinical practice. However, for the conversion rate analyses, there is a separate issue of the reliability of this diagnostic outcome indicator. This metric has drawn criticism for not being able to distinguish reliably between general practices because of the small numbers of cancer diagnosed per practice per year.[193] Future analyses should therefore focus on diagnostic process indicators, such as the age- and sex-adjusted 2WW referral rate, which distinguish between practices reliably.

What this research adds

To our knowledge, there is no comparable literature on the uptake of cancer decision support tools in the UK or elsewhere. Existing qualitative studies shed light on why the uptake of cancer diagnostic tools is relatively low: they suggest that improvements in design and training may improve the uptake of tools, which is in line with our finding of relatively low levels of training in the use of electronic clinical decision support tools.[13,16,19] Indeed, there was some variability within a general practice in terms of its response to the question asking about tool availability. Some GPs in a practice reported that they have access to the tools and others reported that they are unaware of them. This suggests that there is the potential to increase uptake through increased training in use and awareness. Any training should encourage GPs to maximise the amount of information coded into a patient's records. This is because the algorithms rely solely on coded data, and omission of data recorded in text fields is associated with bias.[200]

Conclusion

Our survey indicates that cancer clinical decision support tools are currently not widely used in the UK. This may contribute to our findings in SR1 (see *Chapter 3*) and SR2 (see *Chapter 4*) that there is limited evidence that these tools have an impact on patient outcomes. The upcoming integration of the tools into SystmONE will further increase the potential use of these tools.

As the levels of uptake are relatively low, it remains possible to carry out a RCT to assess whether or not these tools are genuinely helpful in improving the selection of patients for investigation. The attendant benefits of improved selection would include better targeting of investigation resources, potentially earlier diagnosis and reduced treatment costs.[13–15,19,88,89,91,186] Such a trial should include a study of barriers to tool use, and ways to overcome them, as well as a health economics arm.

Chapter 9 Discussion

To our knowledge, this is the first piece of work to bring together evidence on the validation clinical effectiveness, cost-effectiveness, and availability and use of cancer diagnostic tools to aid decision-making in primary care in the UK. To better understand the clinical effectiveness and cost-effectiveness of cancer diagnostic tools to aid decision-making in primary care, SR1, SR2 and the updated review followed prespecified systematic review protocols. Alongside these reviews, a decision model was developed to explore the likely trade-offs in using cancer diagnostic tools in practice, and will be available for updating when new evidence becomes available. Finally, to explore the extent to which GPs currently have access to and use cancer diagnostic tools, a survey of GPs was undertaken. The main findings are discussed in this chapter, as well as gaps in the literature and the strengths and limitations of the research.

What evidence exists on the clinical effectiveness and cost-effectiveness of cancer diagnostic tools in primary care? (Systematic review 1)

Systematic review 1 identified a limited number of heterogeneous studies that provided weak evidence of the impact of these diagnostic tools on patient-related outcomes or referral patterns. This finding is in line with previous reviews across different types of cancers,[40,44] confirming the uncertainty surrounding the clinical utility of diagnostic tools in primary care. Only two RCTs were identified, and the only existing trial of RATs found no statistically significant difference in time to diagnosis between the intervention and control groups. No studies evaluating the cost-effectiveness of these tools was identified, even though the definition of diagnostic tool used in SR1 was extended to include the implementation of any diagnostic model or algorithm.

The results of SR1 are limited by heterogeneity between the studies in terms of the tools evaluated, the populations and the outcomes used. Furthermore, the findings are constrained by the design of the included studies, for example non-randomised designs and infrequent use of patient-reported outcomes.

What diagnostic prediction models exist in the literature that have the potential to be used as tools in primary care? (Systematic review 2)

Eleven site-specific cancer diagnostic prediction models were identified, and covered a range of common and less common cancer sites. These included different versions of the QCancer and RAT models, both developed in the UK. All of the models identified had been developed in Europe.

The risk of bias across studies reporting the development or validation of models was mixed. Notably, the QCancer models were developed using cohort data, and most have been externally validated, but lack impact assessment. In comparison, the RAT models have gathered more evidence of impact on practice, but have been developed from case–control studies, and with limited external validation. With the exception of the CRC RAT model, there were no reports to suggest that models were being updated based on new available data.

What evidence exists on the association between different diagnostic interval durations and patient outcomes? (Updated review)

The findings of a previous systematic review[25] were updated (see *Chapter 5*) to further examine the association between different durations of time from first symptom to diagnosis or treatment and

clinical outcomes across all major cancers. A more in-depth evaluation was conducted of CRC to inform the model.

The literature suggests a U-shaped relationship between diagnostic interval and unfavourable outcomes, which suggests the clinical need to expedite the diagnosis of symptomatic patients to avoid the preventable delays that some patients continue to experience. The review also identified important biases and other factors that may affect the findings of studies in this field.

The majority of the CRC studies found 'no association' between various intervals and patient outcomes. A small number of studies ($n = 4$, although three used the same or overlapping population) reported a positive association between shorter intervals and patient outcomes, but, paradoxically, a small number of studies ($n = 3$) also found a negative association.

What is the likely impact that use of the tools in clinical practice may have on patient outcomes and NHS resources? (Economic model)

There are clear gaps and uncertainties in the evidence base that need to be addressed to fully assess the clinical effectiveness and cost-effectiveness of the tools. There is no available evidence that using the tools has an impact on clinical practice.

To explore trade-offs in the benefits, harms and associated NHS resources related to the use of diagnostic tools in primary care, we developed a simple model of the primary care diagnostic pathway for patients with symptoms suspected to be caused by CRC. Exploration of the model indicates uncertainty as to whether or not the amount of benefit to patients who are identified earlier by diagnostic tools and confirmed as having cancer outweighs the loss in quantity and quality of life in those patients referred for further investigation who do not have cancer. In the absence of good-quality evidence, our model predicts that use of a diagnostic algorithm would extend life expectancy by between 4 days and approximately 2 months among symptomatic patients aged 70 years presenting to their GP with undiagnosed CRC, relative to a hypothetical current practice of management according to results of a FIT alone. Although this amount of benefit may be larger than the loss of life from exposing the overwhelming majority of patients without cancer to the risks of colonoscopy, it is unlikely to justify the additional £25 incurred in health-care costs with the tool per low-risk patient presenting to primary care suspected to have CRC.

The ability to fully evaluate the cost-effectiveness of using diagnostic tools in primary care is restricted because of a lack of evidence on their impact on clinical practice and diagnostic intervals, as well as on current standard practice. In the absence of such evidence, our analysis had to rely on two key assumptions: first, about the relationship between the sensitivity of the test and number of GP visits before referral, which has previously been shown to correlate with time to referral, and, second, about the relationship between the time to referral and time to diagnosis and treatment, so that a reduction in the referral interval was assumed to correspond to an equal reduction in the time to diagnosis and treatment.

Because of the uncertainty behind these assumptions, cost-effectiveness analysis was not the objective of our study. However, we were able to identify that, in addition to research to verify the validity of these assumptions, other key sources of uncertainty in the evidence should be the focus of research, including an evaluation of the tools in comparative studies in low-risk populations and relative to current primary care diagnostic practice. It is noted that the diagnostic accuracy studies should be conducted across the low-risk subpopulation, defined by age and prevalence, and that the sensitivity of standard practice and the specificity of triage tools are two of the most influential parameters for cost-effectiveness, which is also driven by the cost of colonoscopy.

To what extent do general practitioners have access to and use existing cancer decision support tools? (General practitioner survey)

We conducted a UK-wide survey of GPs on the availability of cancer decision support tools. Cancer decision support tools, in either paper or electronic format, were available to GPs in 36.6% (95% CI 30.3% to 43.1%) of UK practices. The proportion of general practices where at least one GP has access to the tools and was likely or very likely to use them was 16.7% (95% CI 12.1% to 22.2%). The proportion of general practices in the UK with access to an electronic clinical decision support tool was 19.0% (95% CI 14.0% to 24.6%).

The survey indicates that cancer clinical decision support tools are currently not widely used in the UK. There was no evidence that access to the tools was associated with a change in the rate of 2WW referrals, or the proportion of those referrals transpiring to have cancer. The subanalyses also did not provide any evidence of an association between access to the tools and practice-level diagnostic activity. This is, perhaps, not surprising, given that the electronic tools, in particular, have been available for only a short time and our survey suggests that they are not yet embedded in clinical practice.

Strengths

Systematic reviews 1 and 2, and the updated review, followed prespecified systematic review protocols, and the team conducting the reviews are independent and experienced in systematic review methodology. The decision model was specifically developed to reflect the primary care pathway for patients with symptoms suspected to be caused by CRC. We were able to represent key relationships required to produce predictions in terms of meaningful clinical outcomes and identify the main parameters contributing to uncertainty in the model. The model is expected to undergo further refinement as new evidence emerges, and to be used as a template for modelling the effect of diagnostic tools in other cancers.

The GP survey is the first UK-wide survey of the availability and use of cancer decision support tools. It is also the first national study to attempt to link use of the tools with changes in the rate of 2WW referrals or cancer diagnoses.

Limitations

The reviews and decision model were limited by a lack of evidence, and poor study quality. The best-quality study in SR1[35] was conducted in Australia, and used a composite intervention including RATs (with older versions of several instruments that were developed using populations from a different country). Not only does this limit the generalisability of results, but it also hinders examination of why no impact of patient outcomes was found. For instance, could it be due to low uptake, poor implementation or limited marginal contribution of the tools in assessing the risk of cancer?

Systematic review 2 identified a number of potential models for use in primary care, yet considerable gaps in the literature exist. There is a disparity in the available evidence on the impact of the tools and the extent to which models have been validated (e.g. whether or not they have been validated at all, and if so, if it has been in the appropriate populations). Comparison of the identified models is limited by heterogeneity between different prediction models for different cancers.

Because there was heterogeneity between studies in the updated review of diagnostic intervals, a 'vote-counting' approach was taken, which does not account for the magnitude of any effect. The potential for selective outcome reporting also limits the interpretation of the evidence on an association between duration of pre-diagnostic intervals and patient outcomes. It is therefore very

difficult to say whether or not there is an association between different interval durations and patient outcomes, although the more recent evidence of a U-shaped relationship requires further examination.

The decision model has several limitations as it relies on a linked-data approach in a context of little available evidence. In particular, in the absence of direct clinical effectiveness data, we assume a relationship between the sensitivity of both the algorithm and routine referral practice and the number of primary care visits before referral, which has been shown to be associated with the referral interval, and that any difference between diagnostic strategies in the implied referral interval equals their difference in time to diagnosis and treatment. The lack of relevant data on the diagnostic accuracy of the current referral practice in primary care, and on the prevalence of cancer in populations at low risk are identified, not surprisingly, to be crucial in calculating the magnitude of the trade-offs in the decision model. Crucially, we do not know how the tools are being used and our main analysis of a diagnostic tool that is used to select patients for direct referral to secondary care and manage the rest by standard investigation may not necessarily be relevant to many localities using the tool. Furthermore, there were insufficient data to evaluate the effect of the tools on quality of life that is mediated by a reduction in the diagnostic interval and associated experience of severe anxiety by patients and their relatives, although our optimistic exploratory analysis revealed that this effect is unlikely to render the tools cost-effective, without further impact on patient survival.

The survey was limited by a lower than expected response rate, with a small likelihood of response bias (those aware and using the tools may have been more likely to respond). This may limit the ability to conclude that the tools do not lead to increased diagnostic activity or, indeed, the ability to characterise referral systems that may be able to benefit more from the tools.

Uncertainties

There are clear gaps and uncertainties in the evidence base that need to be addressed to fully assess the clinical effectiveness and cost-effectiveness of the tools. There is uncertainty about the relative accuracy of the tools, as there is no single study that has compared them, or evaluated any one of them relative to standard practice or against any comparator. The tools that we focused on had evidence from single-arm studies of different designs and population prevalence. Our model-based evaluation was undertaken in the context of the recent publication of a policy recommendation for FITs to replace FOBTs as routine tests. In primary care, the evidence on FITs is limited as the tests have not been evaluated in the population relevant to our study. This means that even the evidence from indirect comparisons of the tools with FITs is limited to diagnostic accuracy data that may not be valid for our study purposes.

Moreover, there is uncertainty about the interaction between GP, tool and the patient. The intervention is not just the tool, but how the GP interacts with the tool. A better understanding of this may help to disentangle uncertainties on whether or not a lack of evidence of the clinical effectiveness of the tools is due to poor implementation, insufficient uptake or the assumptions on which the tools are based.

Given the uncertainities in the evidence base, it is also possible that rapid referral can do more harm than good, or at least more harm than people might expect. As well as the false positives, there is the likelihood that expediting investigation in one group will lead to fewer resources being available for the group that are not expedited. The group not expedited will contain cancer cases, as sensitivity is around 50%, and slower diagnosis could lead to worse outcomes for these cases. These uncertainties and the associated potential for an inappropriate allocation of resources suggests the need for future research to address these issues, as outlined in *Chapter 10, Recommendations for research.*

Given the difficulty of conducting robust clinical studies of diagnostic tools for suspected cancer referrals in primary care, evaluations may pragmatically focus on intermediate outcomes, such as referral intervals, time to diagnosis and treatment initiation. In addressing the structural uncertainty arising from translating an effect on referral intervals into an effect on the time interval until diagnosis, observational studies may fruitfully exploit routine electronic health-care records in which, for example, changes in referral policy guidance could opportunistically be used as a natural experiment to assess the impact of earlier referrals on the diagnostic interval.

Chapter 10 Conclusions

Current evaluations of feature-based cancer diagnostic tools in primary care provided limited and mixed evidence of the impact on patient outcomes. QCancer and RATs are the dominant risk prediction tools identified for use in primary care for any cancer. These tools have the potential to be used by GPs on a wider scale once the knowledge gaps highlighted in the review are addressed. The research has been conducted in the UK, the Netherlands and Australia, suggesting limited geographical application.

The lack of robust clinical effectiveness data is an important limiting factor in assessing the cost-effectiveness of diagnostic tools. Better-quality and more research is needed to provide these data, for example more robust study designs and a wider selection of outcomes. However, identifying the ideal approach may not be straightforward. For example, there is an ongoing debate as to the most appropriate outcome for evaluating interventions to improve cancer diagnosis and referral.[201–203] Moreover, by comparing average times to diagnosis, patients not prioritised for quick referrals are at increased risk of being missed.

The cancer clinical decision support tools are not widely used, and this could explain the limited evidence evaluating their clinical effectiveness. This paucity of evidence, however, makes assessment of the impact of their use on patient outcomes difficult. The levels of uptake of the tools are relatively low, leaving sufficient numbers of general practices to act as controls in future RCTs to assess whether or not these tools are genuinely helpful in improving the selection of patients for investigation.

Recommendations for practice

The uncertainties about the clinical effectiveness of cancer diagnostic tools in primary care, compounded by continued lack of clarity about the relationship between delay to diagnosis or treatment and outcome, suggest that further recommendations for increased use of these tools may be premature. It may be that the limited uptake of the tools observed in our survey is appropriate, given the limited and mixed evidence base we have found.

Recommendations for research

It is undeniable that the rationale for cancer diagnostic tools in primary care is strong. Therefore, the priority should be for further research to confirm or refute the potential of cancer diagnostic tools to improve cancer outcomes. QCancer and RATs are well-established tools. We therefore tentatively suggest that further research should concentrate on QCancer and RATs.

The research on these tools should include:

1. Continued model validation – RATs should try to emulate the validation methods employed by QCancer. Adhering to good conduct and reporting guidance on studies of prediction tools is essential to maximise the value of the studies and to improve the ability to review them.
2. QCancer versus RATs – the performance of the two tools should be compared. It would be particularly helpful to see how similar the risks predicted by the two tools are for identical patients.
3. Assessment of clinical effectiveness – although some assessment of the clinical effectiveness of RATs has been performed, this has been inconclusive, so there is a need for this to be performed for both tools. Studies should collect information on time to diagnosis and treatment and stage at diagnosis, as well as health outcomes (if feasible). In terms of design, cluster randomised trials have been shown to be feasible, and are preferable to quasi-experimental designs already attempted.

Given the likely small effect on health outcomes in particular, it is essential that the sample size of any study is appropriately large. A Phase II cluster RCT evaluating the impact of RATs for gastro-oesophageal cancer is ongoing.[20]

4. Assessment of the tool in GP patient consultations – as the tool is not used in isolation, research to understand the complex nature of the intervention is warranted to investigate the interaction between GP, tool and patient. This may be part of an interventional study.
5. Assessment of quality-of-life implications of delayed and incorrect diagnosis. We do not know what the effect of delayed diagnosis is on the health-related quality of life of cancer patients and their families during the diagnostic delay and after delayed diagnosis. The anxiety caused by persistent unresolved symptoms would require retrospective observational studies recruiting subjects in large secondary care centres.
6. Assessment of cost-effectiveness – although the main current limitation is the absence of clinical effectiveness evidence, researchers should anticipate how cost-effectiveness will be addressed when such evidence becomes available. Incorporating assessment of cost-effectiveness into trials may be one approach, but experience suggests that, to generalise and extrapolate from the trial results, continued enhancement of the economic model developed in this project should be pursued and used to synthesise the latest evidence from points 1 to 3 above. Thus, our model may provide an iterative framework for informing the design of studies described in point 3, and be populated by the results of those studies.
7. Assessment of impact on other referral pathways – clinical effectiveness studies should consider not just the referral/diagnosis time in expedited pathways, but also the impact on the other pathways, as cancer cases might be slowed by a shift in resources to the expedited pathway.
8. Finally, the challenges of interpreting observational data exploring the relationship between delay and outcome continues, as demonstrated in the updated review. For further studies answering this question to be useful, they need to address issues of unmeasured confounding such as grade and stage, as well as problems of survivor time or selection bias implicit in the period of follow-up over which the studies are conducted. These are problems limiting the quality of existing evidence, some of which may be improved by developing agreed standards for providing complete and accurate reporting of all methods used and outcomes measured.

Acknowledgements

We would like to acknowledge the instrumental support of Ms Louise Crathorne and Professor Ruben E Mujica-Mota. We gratefully thank Ms Jenny Lowe for all the support received during this project, and Ms Barbara France for her help with the updated review. We also thank the external reviewers, who provided valuable advice on a complex project. We thank the patients who agreed to participate in a research workshop, as well as the GPs who provided invaluable information for understanding the CRC diagnosis pathway. Our special thanks go to Dr Emma Cockcroft, who undertook the task of organising the meetings with the patients and the GPs.

Contributions of authors

Professor Antonieta Medina-Lara (https://orcid.org/0000-0001-7325-8246), Associate Professor in Health Economics, acted as the project lead, worked on the systematic review of the economic models evidence and on the development of the de novo economic model, and supported the GP survey.

Dr Bogdan Grigore (https://orcid.org/0000-0003-4241-7595), Postdoctoral Research Associate, led SR1 and SR2.

Ms Ruth Lewis (https://orcid.org/0000-0003-0745-995X), Senior Lecturer, led the updated review and worked on SR1 and SR2.

Dr Jaime Peters (https://orcid.org/0000-0003-1778-3518), Senior Research Fellow, supported SR1 and SR2 and the writing-up of the economic model.

Dr Sarah Price (https://orcid.org/0000-0002-2228-2374), Research Fellow, undertook the design, data collection and analysis, and the writing-up of the GP survey.

Dr Paolo Landa (https://orcid.org/0000-0001-6532-6747), Research Fellow in Health Economics, worked on the systematic review of the economic models and on the de novo model.

Ms Sophie Robinson (https://orcid.org/0000-0003-0463-875X), Information Specialist, carried out the electronic searches for SR1, SR2, the review of economic models and the review of parameter values for the de novo economic model.

Professor Richard Neal (https://orcid.org/0000-0002-3544-2744), Professor of Primary Care Oncology, and **Professor William Hamilton (https://orcid.org/0000-0003-1611-1373)**, Professor of Primary Care Diagnostics, provided support to the project on the clinical aspects of the project.

Professor Anne E Spencer (https://orcid.org/0000-0002-8163-3103), Associate Professor in Health Economics, provided senior advice for the review of the economic models, the de novo model and the GP survey.

Data-sharing statement

Requests for survey data or a copy of the model should be submitted to the corresponding author for consideration. Access to anonymised data or a dummy-data populated model may be granted following review.

References

1. De Angelis R, Sant M, Coleman MP, Francisci S, Baili P, Pierannunzio D, *et al.* Cancer survival in Europe 1999-2007 by country and age: results of EUROCARE – 5 – a population-based study. *Lancet Oncol* 2014;**15**:23–34. https://doi.org/10.1016/S1470-2045(13)70546-1
2. 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**:1865–73. https://doi.org/10.1093/ije/dyu174
3. 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**:327–30. https://doi.org/10.5694/mja12.11357
4. Richards MA. The National Awareness and Early Diagnosis Initiative in England: assembling the evidence. *Br J Cancer* 2009;**101**(Suppl. 2):1–4. https://doi.org/10.1038/sj.bjc.6605382
5. 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**:1220–6. https://doi.org/10.1038/bjc.2012.408
6. National Institute for Health and Care Excellence. *Suspected Cancer: Recognition and Referral.* URL: www.nice.org.uk/guidance/ng12 (accessed 16 March 2019).
7. Langton S, Siau D, Bankhead C. Two-week rule in head and neck cancer 2000–14: a systematic review. *Br J Oral Maxillofac Surg* 2016;**54**:120–31. https://doi.org/10.1016/j.bjoms.2015.09.041
8. Hiom SC. Diagnosing cancer earlier: reviewing the evidence for improving cancer survival. *Br J Cancer* 2015;**112**(Suppl. 1):1–5. https://doi.org/10.1038/bjc.2015.23
9. 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**:353–65. https://doi.org/10.1016/S1470-2045(12)70041-4
10. Hendriksen JM, Geersing GJ, Moons KG, de Groot JA. Diagnostic and prognostic prediction models. *J Thromb Haemost* 2013;**11**(Suppl. 1):129–41. https://doi.org/10.1111/jth.12262
11. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, *et al.* Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;**162**:W1–73. https://doi.org/10.7326/M14-0698
12. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, *et al.* Prognosis Research Strategy (PROGRESS) 3: prognostic model research. *PLOS Med* 2013;**10**:e1001381. https://doi.org/10.1371/journal.pmed.1001381
13. 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**:e30–6. https://doi.org/10.3399/bjgp13X660751
14. Collins GS, Altman DG. Identifying patients with undetected pancreatic cancer in primary care: an independent and external validation of QCancer® (pancreas). *Br J Gen Pract* 2013;**63**:e636–42. https://doi.org/10.3399/bjgp13X671623
15. Collins GS, Altman DG. Identifying patients with undetected colorectal cancer: an independent validation of QCancer (colorectal). *Br J Cancer* 2012;**107**:260–5. https://doi.org/10.1038/bjc.2012.266

16. Green T, Martins T, Hamilton W, Rubin G, Elliott K, Macleod U. Exploring GPs' experiences of using diagnostic tools for cancer: a qualitative study in primary care. *Fam Pract* 2015;**32**:101–5. https://doi.org/10.1093/fampra/cmu081

17. 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. https://doi.org/10.1017/S1463423615000109

18. Chiang PP, 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):77–83. https://doi.org/10.1038/bjc.2015.46

19. Moffat J, Ironmonger L, Green T. *Clinical Decision Support Tool for Cancer (CDS) Project: Evaluation Report to the Department of Health*. 2014. URL: www.cancerresearchuk.org/sites/default/files/cds_final_310714.pdf (accessed January 2020).

20. Moore HJ, Nixon C, Tariq A, Emery J, Hamilton W, Hoare Z, *et al.* Evaluating a computer aid for assessing stomach symptoms (ECASS): study protocol for a randomised controlled trial. *Trials* 2016;**17**:184. https://doi.org/10.1186/s13063-016-1307-3

21. Iyen-Omofoman B, Tata LJ, Baldwin DR, Smith CJ, Hubbard RB. Using socio-demographic and early clinical features in general practice to identify people with lung cancer earlier. *Thorax* 2013;**68**:451–9. https://doi.org/10.1136/thoraxjnl-2012-202348

22. 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. *Gut* 2011;**60**:1242–8. https://doi.org/10.1136/gut.2010.225987

23. Bourne P, Rosendahl C, Keir J, Cameron A. BLINCK-A diagnostic algorithm for skin cancer diagnosis combining clinical features with dermatoscopy findings. *Dermatol Pract Concept* 2012;**2**:202a12. https://doi.org/10.5826/dpc.0202a12

24. Goff BA, Mandel LS, Melancon CH, Muntz HG. Frequency of symptoms of ovarian cancer in women presenting to primary care clinics. *JAMA* 2004;**291**:2705–12. https://doi.org/10.1001/jama.291.22.2705

25. Neal RD, Tharmanathan P, France B, Din NU, Cotton S, Fallon-Ferguson J, *et al.* Is increased time to diagnosis and treatment in symptomatic cancer associated with poorer outcomes? Systematic review. *Br J Cancer* 2015;**112**(Suppl. 1):92–107. https://doi.org/10.1038/bjc.2015.48

26. Bouwmeester W, Zuithoff NP, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, *et al.* Reporting and methods in clinical prediction research: a systematic review. *PLOS Med* 2012;**9**:1–12. https://doi.org/10.1371/journal.pmed.1001221

27. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Ann Intern Med* 2006;**144**:201–9. https://doi.org/10.7326/0003-4819-144-3-200602070-00009

28. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ* 2009;**338**:b606. https://doi.org/10.1136/bmj.b606

29. Collins GS, de Groot JA, Dutton S, Omar O, Shanyinde M, Tajar A, *et al.* External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol* 2014;**14**:40. https://doi.org/10.1186/1471-2288-14-40

30. Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, *et al.* Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLOS Med* 2014;**11**:e1001744. https://doi.org/10.1371/journal.pmed.1001744

31. Centre for Reviews and Dissemination. *Systematic Reviews: CRD's Guidance for Undertaking Reviews in Health Care*. 2009. URL: www.york.ac.uk/media/crd/Systematic_Reviews.pdf (accessed January 2020).
32. 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**:e38–45. https://doi.org/10.3399/bjgp12X616355
33. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009;**339**:b2535. https://doi.org/10.1136/bmj.b2535
34. Higgins JPT, Altman DG, Sterne JAC. Chapter 8: Assessing Risk of Bias in Included Studies. In Higgins JPT, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0*. (updated March 2011). Chichester: John Wiley & Sons; 2011.
35. Emery JD, Gray V, Walter FM, Cheetham S, Croager EJ, Slevin T, *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**:1459–69. https://doi.org/10.1038/bjc.2017.310
36. English DR, Burton RC, del Mar CB, Donovan RJ, Ireland PD, Emery G. Evaluation of aid to diagnosis of pigmented skin lesions in general practice: controlled trial randomised by practice. *BMJ* 2003;**327**:375. https://doi.org/10.1136/bmj.327.7411.375
37. Del Mar CB, Green AC. Aid to diagnosis of melanoma in primary medical care. *BMJ* 1995;**310**:492–5. https://doi.org/10.1136/bmj.310.6978.492
38. Gulati A, Harwood CA, Rolph J, Pottinger E, McGregor JM, Goad N, Proby CM. Is an online skin cancer toolkit an effective way to educate primary care physicians about skin cancer diagnosis and referral? *J Eur Acad Dermatol Venereol* 2015;**29**:2152–9. https://doi.org/10.1111/jdv.13167
39. 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**(Suppl. 2):80–6. https://doi.org/10.1038/sj.bjc.6605396
40. Williams TG, 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**:63. https://doi.org/10.1186/s12876-016-0475-7
41. 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**:e396–e404. https://doi.org/10.3399/bjgp17X690917
42. Usher-Smith J, Emery J, Hamilton W, Griffin SJ, Walter FM. Risk prediction tools for cancer in primary care. *Br J Cancer* 2015;**113**:1645–50. https://doi.org/10.1038/bjc.2015.409
43. Walker JG, Licqurish S, Chiang PP, Pirotta M, Emery JD. Cancer risk assessment tools in primary care: a systematic review of randomized controlled trials. *Ann Fam Med* 2015;**13**:480–9. https://doi.org/10.1370/afm.1837
44. Usher-Smith JA, Walter FM, Emery JD, Win AK, Griffin SJ. Risk prediction models for colorectal cancer: a systematic review. *Cancer Prev Res* 2016;**9**:13–26. https://doi.org/10.1158/1940-6207.CAPR-15-0274
45. Adams ST, Leveson SH. Clinical prediction rules. *BMJ* 2012;**344**:d8312. https://doi.org/10.1136/bmj.d8312

46. Wolff R, Whiting P, Mallett S, Riley R, Westwood M, Kleijnen J, Moons K. PROBAST: A Risk of Bias Tool for Prediction Modelling Studies. In *Filtering the Information Overload for Better Decisions*. Abstracts of the 23rd Cochrane Colloquium, 3–7 October 2015, Vienna, Austria.

47. Debray TP, Damen JA, Snell KI, Ensor J, Hooft L, Reitsma JB, *et al.* A guide to systematic review and meta-analysis of prediction model performance. *BMJ* 2017;**356**:i6460. https://doi.org/10.1136/bmj.i6460

48. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, *et al.* PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med* 2019;**170**:51–8. https://doi.org/10.7326/M18-1376

49. Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, *et al.* PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Intern Med* 2019;**170**:W1–W33. https://doi.org/10.7326/M18-1377

50. Elias SG, Kok L, Witteman BJM, Goedhard JG, Romberg-Camps MJL, Muris JWM, *et al.* Published diagnostic models safely excluded colorectal cancer in an independent primary care validation study. *J Clin Epidemiol* 2017;**82**:149–57.e8. https://doi.org/10.1016/j.jclinepi.2016.09.014

51. Hamilton W, Peters TJ, Round A, Sharp D. What are the clinical features of lung cancer before the diagnosis is made? A population based case–control study. *Thorax* 2005;**60**:1059–65. https://doi.org/10.1136/thx.2005.045880

52. Hippisley-Cox J, Coupland C. Identifying patients with suspected lung cancer in primary care: derivation and validation of an algorithm. *Br J Gen Pract* 2011;**61**:e715–23. https://doi.org/10.3399/bjgp11X606627

53. 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. https://doi.org/10.3399/bjgp13X660733

54. 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. https://doi.org/10.3399/bjgp13X660724

55. 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**:279–86. https://doi.org/10.1093/fampra/12.3.279

56. 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**:313–16.

57. 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**:160–5. https://doi.org/10.1093/fampra/13.2.160

58. Jones R, Latinovic R, Charlton J, Gulliford MC. Alarm symptoms in early diagnosis of cancer in primary care: cohort study using General Practice Research Database. *BMJ* 2007;**334**:1040. https://doi.org/10.1136/bmj.39171.637106.AE

59. Jordan KP, Hayward RA, Blagojevic-Bucknall M, Croft P. Incidence of prostate, breast, lung and colorectal cancer following new consultation for musculoskeletal pain: a cohort study among UK primary care patients. *Int J Cancer* 2013;**133**:713–20. https://doi.org/10.1002/ijc.28055

60. McCowan C, Donnan PT, Dewar J, Thompson A, Fahey T. Identifying suspected breast cancer: development and validation of a clinical prediction rule. *Br J Gen Pract* 2011;**61**:e205–14. https://doi.org/10.3399/bjgp11X572391

61. Kop R, Hoogendoorn M, Moons LMG, Numans ME, ten Teije A. On the advantage of using dedicated data mining techniques to predict colorectal cancer. *Lect Notes Comp Sci* 2015;**9105**:133–42. https://doi.org/10.1007/978-3-319-19551-3_16

62. Keane MG, Horsfall L, Rait G, Pereira SP. A case–control study comparing the incidence of early symptoms in pancreatic and biliary tract cancer. *BMJ Open* 2014;**4**:e005720. https://doi.org/10.1136/bmjopen-2014-005720

63. Holtedahl K, Hjertholm P, Borgquist L, Donker GA, Buntinx F, Weller D, *et al.* Abdominal symptoms and cancer in the abdomen: prospective cohort study in European primary care. *Br J Gen Pract* 2018;**68**:e301–e310. https://doi.org/10.3399/bjgp18X695777

64. Dommett RM, Redaniel T, Stevens MC, Martin RM, Hamilton W. Risk of childhood cancer with symptoms in primary care: a population-based case–control study. *Br J Gen Pract* 2013;**63**:e22–9. https://doi.org/10.3399/bjgp13X660742

65. Hamilton W. Cancer diagnosis in primary care. *Br J Gen Pract* 2010;**60**:121–8. https://doi.org/10.3399/bjgp10X483175

66. 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**:e250–5. https://doi.org/10.3399/bjgp13X665215

67. Shephard EA, Neal RD, Rose PW, Walter FM, Hamilton W. Symptoms of adult chronic and acute leukaemia before diagnosis: large primary care case-control studies using electronic records. *Br J Gen Pract* 2016;**66**:e182–8. https://doi.org/10.3399/bjgp16X683989

68. 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**:e106–13. https://doi.org/10.3399/bjgp15X683545

69. 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**:e598–604. https://doi.org/10.3399/bjgp12X654560

70. 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**:1940–4. https://doi.org/10.1038/bjc.2012.190

71. 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**:25–31. https://doi.org/10.1038/bjc.2012.551

72. Stapley SA, Rubin GP, Alsina D, Shepard 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**:e336–44. https://doi.org/10.3399/bjgp17X690425

73. 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**:e643–8. https://doi.org/10.3399/bjgp13X671632

74. 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. https://doi.org/10.1186/1741-7015-7-17

75. 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**:323–7. https://doi.org/10.1038/sj.bjc.6604165

76. 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**:393–401. https://doi.org/10.1002/pds.1335

77. Hodder RJ, Ballal M, Selvachandran SN, Cade D. Pitfalls in the construction of cancer guidelines demonstrated by the analyses of colorectal referrals. *Ann R Coll Surg Engl* 2005;**87**:419–26. https://doi.org/10.1308/003588405X71018

78. 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* 2016;**69**:53–61. https://doi.org/10.1016/j.artmed.2016.03.003

79. Kop R, Hoogendoorn M, Teije AT, Büchner FL, Slottje P, Moons LM, 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. https://doi.org/10.1016/j.compbiomed.2016.06.019

80. 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**:e29–37. https://doi.org/10.3399/bjgp12X616346

81. 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**:399–405. https://doi.org/10.1038/sj.bjc.6602714

82. 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. https://doi.org/10.3399/bjgp11X606609

83. Hamilton W, Peters TJ, Bankhead C, Sharp D. Risk of ovarian cancer in women with symptoms in primary care: population based case–control study. *BMJ* 2009;**339**:b2998. https://doi.org/10.1136/bmj.b2998

84. Hippisley-Cox J, Coupland C. Identifying women with suspected ovarian cancer in primary care: derivation and validation of algorithm. *BMJ* 2011;**344**:d8009. https://doi.org/10.1136/bmj.d8009

85. Hippisley-Cox J, Coupland C. Identifying patients with suspected renal tract cancer in primary care: derivation and validation of an algorithm. *Br J Gen Pract* 2012;**62**:e251–60. https://doi.org/10.3399/bjgp12X636074

86. 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**:e516–22. https://doi.org/10.3399/bjgp15X686077

87. Hamilton W, Kernick D. Clinical features of primary brain tumours: a case–control study using electronic primary care records. *Br J Gen Pract* 2007;**57**:695–9.

88. Collins GS, Altman DG. Identifying patients with undetected gastro-oesophageal cancer in primary care: external validation of QCancer® (gastro-oesophageal). *Eur J Cancer* 2013;**49**:1040–8. https://doi.org/10.1016/j.ejca.2012.10.023

89. Collins GS, Altman DG. Identifying women with undetected ovarian cancer: independent and external validation of QCancer® (ovarian) prediction model. *Eur J Cancer Care* 2013;**22**:423–9. https://doi.org/10.1111/ecc.12015

90. Hamilton W, Sharp DJ, Peters TJ, Round AP. Clinical features of prostate cancer before diagnosis: a population-based, case-control study. *Br J Gen Pract* 2006;**56**:756–62.

91. Collins GS, Altman DG. Identifying patients with undetected renal tract cancer in primary care: an independent and external validation of QCancer® (renal) prediction model. *Cancer Epidemiol* 2013;**37**:115–20. https://doi.org/10.1016/j.canep.2012.11.005

92. Weller D, Vedsted P, Rubin G, Walter FM, Emery J, Scott S, *et al.* The Aarhus statement: improving design and reporting of studies on early cancer diagnosis. *Br J Cancer* 2012;**106**:1262–7. https://doi.org/10.1038/bjc.2012.68

93. 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**:1119–26. https://doi.org/10.1016/S0140-6736(99)02143-1

94. Ramos M, Esteva M, Cabeza E, Campillo C, Llobera J, Aguiló A. Relationship of diagnostic and therapeutic delay with survival in colorectal cancer: a review. *Eur J Cancer* 2007;**43**:2467–78. https://doi.org/10.1016/j.ejca.2007.08.023

95. Ramos M, Esteva M, Cabeza E, Llobera J, Ruiz A. Lack of association between diagnostic and therapeutic delay and stage of colorectal cancer. *Eur J Cancer* 2008;**44**:510–21. https://doi.org/10.1016/j.ejca.2008.01.011

96. Thompson MR, Heath I, Swarbrick ET, Faulds Wood L, Ellis BG. Earlier diagnosis and treatment of symptomatic bowel cancer: can it be achieved and how much will it improve survival? *Colorectal Dis* 2011;**13**:6–16. https://doi.org/10.1111/j.1463-1318.2009.01986.x

97. Moher D, Tsertsvadze A, Tricco AC, Eccles M, Grimshaw J, Sampson M, Barrowman N. When and how to update systematic reviews. *Cochrane Database Syst Rev* 2008;**1**:MR000023. https://doi.org/10.1002/14651858.MR000023.pub3

98. Cooper C, Booth A, Britten N, Garside R. A comparison of results of empirical studies of supplementary search techniques and recommendations in review methodology handbooks: a methodological review. *Syst Rev* 2017;**6**:234. https://doi.org/10.1186/s13643-017-0625-1

99. Tørring ML, Frydenberg M, Hansen RP, Olesen F, Vedsted P. Evidence of increasing mortality with longer diagnostic intervals for five common cancers: a cohort study in primary care. *Eur J Cancer* 2013;**49**:2187–98. https://doi.org/10.1016/j.ejca.2013.01.025

100. Ángeles-Llerenas A, Torres-Mejía G, Lazcano-Ponce E, Uscanga-Sánchez S, Mainero-Ratchelous F, Hernández-Ávila JE, *et al.* Effect of care-delivery delay on the survival of Mexican women with breast cancer. *Salud Publica Mex* 2016;**58**:237–50. https://doi.org/10.21149/spm.v58i2.7793

101. Redaniel MT, Martin RM, Ridd MJ, Wade J, Jeffreys M. Diagnostic intervals and its association with breast, prostate, lung and colorectal cancer survival in England: historical cohort study using the Clinical Practice Research Datalink. *PLOS ONE* 2015;**10**:e0126608. https://doi.org/10.1371/journal.pone.0126608

102. Murchie P, Raja EA, Lee AJ, Brewster DH, Campbell NC, Gray NM, *et al.* Effect of longer health service provider delays on stage at diagnosis and mortality in symptomatic breast cancer. *Breast* 2015;**24**:248–55. https://doi.org/10.1016/j.breast.2015.02.027

103. Pace LE, Mpunga T, Hategekimana V, Dusengimana JM, Habineza H, Bigirimana JB, *et al.* Delays in breast cancer presentation and diagnosis at two rural cancer referral centers in Rwanda. *Oncologist* 2015;**20**:780–8. https://doi.org/10.1634/theoncologist.2014-0493

104. Unger-Saldaña K, Miranda A, Zarco-Espinosa G, Mainero-Ratchelous F, Bargalló-Rocha E, Miguel Lázaro-León J. Health system delay and its effect on clinical stage of breast cancer: multicenter study. *Cancer* 2015;**121**:2198–206. https://doi.org/10.1002/cncr.29331

105. Bryan RT. Cell adhesion and urothelial bladder cancer: the role of cadherin switching and related phenomena. *Philos Trans R Soc B* 2015;**370**:20140042. https://doi.org/10.1098/rstb.2014.0042

106. Helewa RM, Turner D, Park J, Wirtzfeld D, Czaykowski P, Hochman D, *et al.* Longer waiting times for patients undergoing colorectal cancer surgery are not associated with decreased survival. *J Surg Oncol* 2013;**108**:378–84. https://doi.org/10.1002/jso.23412

107. Murchie P, Raja EA, Brewster DH, Campbell NC, Ritchie LD, Robertson R, *et al.* Time from first presentation in primary care to treatment of symptomatic colorectal cancer: effect on disease stage and survival. *Br J Cancer* 2014;**111**:461–9. https://doi.org/10.1038/bjc.2014.352

108. Dregan A, Møller H, Charlton J, Gulliford MC. Are alarm symptoms predictive of cancer survival?: population-based cohort study. *Br J Gen Pract* 2013;**63**:e807–12. https://doi.org/10.3399/bjgp13X675197

109. Pita-Fernández S, González-Sáez L, López-Calviño B, Seoane-Pillado T, Rodríguez-Camacho E, Pazos-Sierra A, *et al.* Effect of diagnostic delay on survival in patients with colorectal cancer: a retrospective cohort study. *BMC Cancer* 2016;**16**:664. https://doi.org/10.1186/s12885-016-2717-z

110. Aslam MI, Chaudhri S, Singh B, Jameson JS. The 'two-week wait' referral pathway is not associated with improved survival for patients with colorectal cancer. *Int J Surg* 2017;**43**:181–5. https://doi.org/10.1016/j.ijsu.2017.05.046

111. Patel R, Anderson JE, McKenzie C, Simpson M, Singh N, Ruzvidzo F, *et al.* Compliance with the 62-day target does not improve long-term survival. *Int J Colorectal Dis* 2018;**33**:65–9. https://doi.org/10.1007/s00384-017-2930-5

112. Chen FW, Sundaram V, Chew TA, Ladabaum U. Advanced-stage colorectal cancer in persons younger than 50 years not associated with longer duration of symptoms or time to diagnosis. *Clin Gastroenterol Hepatol* 2017;**15**:728–37.e3. https://doi.org/10.1016/j.cgh.2016.10.038

113. Leiva A, Esteva M, Llobera J, Macià F, Pita-Fernández S, González-Luján L, *et al.* Time to diagnosis and stage of symptomatic colorectal cancer determined by three different sources of information: a population based retrospective study. *Cancer Epidemiol* 2017;**47**:48–55. https://doi.org/10.1016/j.canep.2016.10.021

114. Janssen RM, Takach O, Nap-Hill E, Enns RA. Time to endoscopy in patients with colorectal cancer: analysis of wait-times. *Can J Gastroenterol Hepatol* 2016;**2016**:8714587. https://doi.org/10.1155/2016/8714587

115. Radzikowska E, Roszkowski-Sliz K, Chabowski M, Glaz P. Influence of delays in diagnosis and treatment on survival in small cell lung cancer patients. *Adv Exp Med Biol* 2013;**788**:355–62. https://doi.org/10.1007/978-94-007-6627-3_48

116. Živković D. Effect of delays on survival in patients with lung carcinoma in Montenegro. *Acta Clin Croat* 2014;**53**:390–8.

117. Gonzalez-Barcala FJ, Falagan JA, Garcia-Prim JM, Valdes L, Carreira JM, Pose A, *et al.* Symptoms and reason for a medical visit in lung cancer patients. *Acta Med Port* 2014;**27**:318–24. https://doi.org/10.20344/amp.4229

118. Gildea TR, DaCosta Byfield S, Hogarth DK, Wilson DS, Quinn CC. A retrospective analysis of delays in the diagnosis of lung cancer and associated costs. *Clinicoecon Outcomes Res* 2017;**9**:261–9. https://doi.org/10.2147/CEOR.S132259

119. Kim JO, Davis F, Butts C, Winget M. Waiting time intervals for non-small cell lung cancer diagnosis and treatment in alberta: quantification of intervals and identification of risk factors associated with delays. *Clin Oncol* 2016;**28**:750–9. https://doi.org/10.1016/j.clon.2016.06.010

120. Nikonova A, Guirguis HR, Buckstein R, Cheung MC. Predictors of delay in diagnosis and treatment in diffuse large B-cell lymphoma and impact on survival. *Br J Haematol* 2015;**168**:492–500. https://doi.org/10.1111/bjh.13150

121. Goldschmidt N, Zamir L, Poperno A, Kahan NR, Paltiel O. Presenting signs of multiple myeloma and the effect of diagnostic delay on the prognosis. *J Am Board Fam Med* 2016;**29**:702–9. https://doi.org/10.3122/jabfm.2016.06.150393

122. Keizer AC, Korse TCM, Meijer WG, Tesselaar ET. The effect of delay in diagnosis in patients with neuroendocrine tumors. *Int J Endocr Oncol* 2016;**3**:33–9. https://doi.org/10.2217/ije.15.32

123. Alahapperuma LS, Fernando EA. Patient-linked factors associated with delayed reporting of oral and pharyngeal carcinoma among patients attending national cancer institute, Maharagama, Sri Lanka. *Asian Pac J Cancer Prev* 2017;**18**:321–5.

124. Esmaelbeigi F, Hadji M, Harirchi I, Omranipour R, vand Rajabpour M, Zendehdel K. Factors affecting professional delay in diagnosis and treatment of oral cancer in Iran. *Arch Iran Med* 2014;**17**:253–7. https://doi.org/014174/AIM.007

125. Lim A, Mesher D, Gentry-Maharaj A, Balogun N, Widschwendter M, Jacobs I, *et al.* Time to diagnosis of type I or II invasive epithelial ovarian cancers: a multicentre observational study using patient questionnaire and primary care records. *BJOG* 2016;**123**:1012–20. https://doi.org/10.1111/1471-0528.13447

126. Altman AD, Lambert P, Love AJ, Turner D, Lotocki R, Dean E, *et al.* Examining the effects of time to diagnosis, income, symptoms, and incidental detection on overall survival in epithelial ovarian cancer: Manitoba Ovarian Cancer Outcomes (MOCO) study group. *Int J Gynecol Cancer* 2017;**27**:1637–44. https://doi.org/10.1097/IGC.0000000000001074

127. Gobbi PG, Bergonzi M, Comelli M, Villano L, Pozzoli D, Vanoli A, Dionigi P. The prognostic role of time to diagnosis and presenting symptoms in patients with pancreatic cancer. *Cancer Epidemiol* 2013;**37**:186–90. https://doi.org/10.1016/j.canep.2012.12.002

128. Jooste V, Dejardin O, Bouvier V, Arveux P, Maynadie M, Launoy G, Bouvier AM. Pancreatic cancer: wait times from presentation to treatment and survival in a population-based study. *Int J Cancer* 2016;**139**:1073–80. https://doi.org/10.1002/ijc.30166

129. Gao W, Song LB, Yang J, Song NH, Wu XF, Song NJ, *et al.* Risk factors and negative consequences of patient's delay for penile carcinoma. *World J Surg Oncol* 2016;**14**:124. https://doi.org/10.1186/s12957-016-0863-z

130. Bonfill X, Martinez-Zapata MJ, Vernooij RW, Sánchez MJ, Suárez-Varela MM, de la Cruz J, *et al.* Clinical intervals and diagnostic characteristics in a cohort of prostate cancer patients in Spain: a multicentre observational study. *BMC Urol* 2015;**15**:60. https://doi.org/10.1186/s12894-015-0058-x

131. Urakawa H, Tsukushi S, Arai E, Kozawa E, Futamura N, Ishiguro N, Nishida Y. Association of short duration from initial symptoms to specialist consultation with poor survival in soft-tissue sarcomas. *Am J Clin Oncol* 2015;**38**:266–71. https://doi.org/10.1097/COC.0b013e318295aea2

132. Goedhart LM, Gerbers JG, Ploegmakers JJ, Jutte PC. Delay in diagnosis and its effect on clinical outcome in high-grade sarcoma of bone: a referral oncological centre study. *Orthop Surg* 2016;**8**:122–8. https://doi.org/10.1111/os.12239

133. De Boer C, Niyonzima N, Orem J, Bartlett J, Zafar SY. Prognosis and delay of diagnosis among Kaposi's sarcoma patients in Uganda: a cross-sectional study. *Infect Agent Cancer* 2014;**9**:17. https://doi.org/10.1186/1750-9378-9-17

134. Kobayashi K, Saito T, Kitamura Y, Nobushita T, Kawasaki T, Hara N, Takahashi K. Effect of the time from the presentation of symptoms to medical consultation on primary tumor size and survival in patients with testicular cancer: shift in the last 2 decades. *Urol Oncol* 2014;**32**:43.e17–22. https://doi.org/10.1016/j.urolonc.2013.05.007

135. Tørring ML, Frydenberg M, Hamilton W, Hansen RP, Lautrup MD, Vedsted P. Diagnostic interval and mortality in colorectal cancer: U-shaped association demonstrated for three different datasets. *J Clin Epidemiol* 2012;**65**:669–78. https://doi.org/10.1016/j.jclinepi.2011.12.006

136. Tørring ML, Frydenberg M, Hansen RP, Olesen F, Hamilton W, Vedsted P. Time to diagnosis and mortality in colorectal cancer: a cohort study in primary care. *Br J Cancer* 2011;**104**:934–40. https://doi.org/10.1038/bjc.2011.60

137. Pruitt SL, Harzke AJ, Davidson NO, Schootman M. Do diagnostic and treatment delays for colorectal cancer increase risk of death? *Cancer Causes Control* 2013;**24**:961–77. https://doi.org/10.1007/s10552-013-0172-6

138. Korsgaard M, Pedersen L, Sørensen HT, Laurberg S. Reported symptoms, diagnostic delay and stage of colorectal cancer: a population-based study in Denmark. *Colorectal Dis* 2006;**8**:688–95. https://doi.org/10.1111/j.1463-1318.2006.01014.x

139. Biswas M, Ades AE, Hamilton W. Symptom lead times in lung and colorectal cancers: what are the benefits of symptom-based approaches to early diagnosis? *Br J Cancer* 2015;**112**:271–7. https://doi.org/10.1038/bjc.2014.597

140. National Institute for Health and Care Excellence. *Quantitative Faecal Immunochemical Tests to Guide Referral for Colorectal Cancer in Primary Care. Diagnostics Guidance [DG30]*. URL: www.nice.org.uk/guidance/dg30 (accessed 16 March 2019).

141. Hamilton W. *Towards Earlier Diagnosis of Cancer in Primary Care: A Population-based Case–Control Study of Colorectal, Lung and Prostate Cancer*. PhD thesis. Bristol: Bristol University; 2005.

142. Caro JJ, Briggs AH, Siebert U, Kuntz KM, ISPOR-SMDM Modeling Good Research Practices Task Force. Modeling good research practices – overview: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force – 1. *Value Health* 2012;**15**:796–803. https://doi.org/10.1016/j.jval.2012.06.012

143. Kaltenthaler E, Tappenden P, Paisley S, Squires H. *NICE DSU Technical Support Document 13: Identifying and Reviewing Evidence to Inform the Conceptualisation and Population of Cost-effectiveness Models*. Sheffield: Decision Support Unit, School of Health and Related Research, University of Sheffield; 2011.

144. Allen E, Nicolaidis C, Helfand M. The evaluation of rectal bleeding in adults. A cost-effectiveness analysis comparing four diagnostic strategies. *J Gen Intern Med* 2005;**20**:81–90. https://doi.org/10.1111/j.1525-1497.2005.40077.x

145. Tappenden P, Chilcott J, Eggington S, Patnick J, Sakai H, Karnon J. Option appraisal of population-based colorectal cancer screening programmes in England. *Gut* 2007;**56**:677–84. https://doi.org/10.1136/gut.2006.095109

146. Tsoi KK, Ng SS, Leung MC, Sung JJ. Cost-effectiveness analysis on screening for colorectal neoplasm and management of colorectal cancer in Asia. *Aliment Pharmacol Ther* 2008;**28**:353–63. https://doi.org/10.1111/j.1365-2036.2008.03726.x

147. Zauber AG, Lansdorp-Vogelaar I, Knudsen AB, Wilschut J, van Ballegooijen M, Kuntz KM. Evaluating test strategies for colorectal cancer screening: a decision analysis for the U.S. Preventive Services Task Force. *Ann Intern Med* 2008;**149**:659–69. https://doi.org/10.7326/0003-4819-149-9-200811040-00244

148. Heitman SJ, Hilsden RJ, Au F, Dowden S, Manns BJ. Colorectal cancer screening for average-risk North Americans: an economic evaluation. *PLOS Med* 2010;**7**:e1000370. https://doi.org/10.1371/journal.pmed.1000370

149. Lee D, Muston D, Sweet A, Cunningham C, Slater A, Lock K. Cost effectiveness of CT colonography for UK NHS colorectal cancer screening of asymptomatic adults aged 60–69 years. *Appl Health Econ Health Policy* 2010;**8**:141–54. https://doi.org/10.2165/11535650-000000000-00000

150. Knudsen AB, Hur C, Gazelle GS, Schrag D, McFarland EG, Kuntz KM. Rescreening of persons with a negative colonoscopy result: results from a microsimulation model. *Ann Intern Med* 2012;**157**:611–20. https://doi.org/10.7326/0003-4819-157-9-201211060-00005

151. Sharp L, Tilson L, Whyte S, O'Ceilleachair A, Walsh C, Usher C, *et al.* Cost-effectiveness of population-based screening for colorectal cancer: a comparison of guaiac-based faecal occult blood testing, faecal immunochemical testing and flexible sigmoidoscopy. *Br J Cancer* 2012;**106**:805–16. https://doi.org/10.1038/bjc.2011.580

152. Whyte S, Chilcott J, Halloran S. Reappraisal of the options for colorectal cancer screening in England. *Colorectal Dis* 2012;**14**:e547–61. https://doi.org/10.1111/j.1463-1318.2012.03014.x

153. Goede SL, van Roon AH, Reijerink JC, van Vuuren AJ, Lansdorp-Vogelaar I, Habbema JD, *et al.* Cost-effectiveness of one versus two sample faecal immunochemical testing for colorectal cancer screening. *Gut* 2013;**62**:727–34. https://doi.org/10.1136/gutjnl-2011-301917

154. Gomes M, Aldridge RW, Wylie P, Bell J, Epstein O. Cost-effectiveness analysis of 3-D computerized tomography colonography versus optical colonoscopy for imaging symptomatic gastroenterology patients. *Appl Health Econ Health Policy* 2013;**11**:107–17. https://doi.org/10.1007/s40258-013-0019-z

155. Tappenden P, Chilcott J, Brennan A, Squires H, Glynne-Jones R, Tappenden J. Using whole disease modeling to inform resource allocation decisions: economic evaluation of a clinical guideline for colorectal cancer using a single model. *Value Health* 2013;**16**:542–53. https://doi.org/10.1016/j.jval.2013.02.012

156. Whyte S, Harnan S. Effectiveness and cost-effectiveness of an awareness campaign for colorectal cancer: a mathematical modeling study. *Cancer Causes Control* 2014;**25**:647–58. https://doi.org/10.1007/s10552-014-0366-6

157. Cantor SB, Rajan T, Linder SK, Volk RJ. A framework for evaluating the cost-effectiveness of patient decision aids: a case study using colorectal cancer screening. *Prev Med* 2015;**77**:168–73. https://doi.org/10.1016/j.ypmed.2015.05.003

158. Pil L, Fobelets M, Putman K, Trybou J, Annemans L. Cost-effectiveness and budget impact analysis of a population-based screening program for colorectal cancer. *Eur J Intern Med* 2016;**32**:72–8. https://doi.org/10.1016/j.ejim.2016.03.031

159. Wong MC, Ching JY, Chan VC, Lam TY, Luk AK, Wong SH, *et al.* Colorectal cancer screening based on age and gender: a cost-effectiveness analysis. *Medicine* 2016;**95**:e2739. https://doi.org/10.1097/MD.0000000000002739

160. Coldman A, Flanagan W, Nadeau C, Wolfson M, Fitzgerald N, Mermon S, *et al.* Projected effect of fecal immunochemical test threshold for colorectal cancer screening on outcomes and costs for Canada using the OncoSim microsimulation model. *J Cancer Policy* 2017;**13**:38–46. https://doi.org/10.1016/j.jcpo.2017.07.004

161. Westwood M, Lang S, Armstrong N, van Turenhout S, Cubiella J, Stirk L, *et al.* Faecal immunochemical tests (FIT) can help to rule out colorectal cancer in patients presenting in primary care with lower abdominal symptoms: a systematic review conducted to inform new NICE DG30 diagnostic guidance. *BMC Med* 2017;**15**:189. https://doi.org/10.1186/s12916-017-0944-z

162. Philips Z, Bojke L, Sculpher M, Claxton K, Golder S. Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. *PharmacoEconomics* 2006;**24**:355–71. https://doi.org/10.2165/00019053-200624040-00006

163. Tappenden P, Eggington S, Nixon J *et al. Colorectal Cancer Screening Options Appraisal. Report to the English Bowel Cancer Screening Working Group*. Sheffield: 2004.

164. Office of National Statistics. *Mortality statistics: Cause [series DH2, no. 32]*. Newport: Office for National Statistics; 2006 [online]. URL: www.statistics.gov.uk/downloads/theme_health/Dh2_32/DH2_No32_2005.pdf (accessed 22 March 2010).

165. Hospital Authority. *Hong Kong Cancer Registry*. URL: www.ha.org.hk/cancereg (accessed May 2007).

166. Claxton K, Sculpher M, Drummond M. A rational framework for decision making by the National Institute For Clinical Excellence (NICE). *Lancet* 2002;**360**:711–15. https://doi.org/10.1016/S0140-6736(02)09832-X

167. Hoomans T, Seidenfeld J, Basu A, Meltzer D. *Systematizing the Use of Value of Information Analysis in Prioritizing Systematic Reviews*. Rockville, MD: Agency for Healthcare Research and Quality; 2012.

168. Ades AE, Lu G, Claxton K. Expected value of sample information calculations in medical decision modeling. *Med Decis Making* 2004;**24**:207–27. https://doi.org/10.1177/0272989X04263162

169. Cancer Research UK. *Cancer Incidence for Common Cancers*. URL: www.cancerresearchuk.org/health-professional/cancer-statistics/incidence/common-cancers-compared (accessed 5 May 2016).

170. Goodwin E, Boddy K, Tatnell L, Hawton A. Involving members of the public in health economics research: insights from selecting health states for valuation to estimate quality-adjusted life-year (QALY) weights. *Appl Health Econ Health Policy* 2018;**16**:187–94. https://doi.org/10.1007/s40258-017-0355-5

171. Pizzo E, Doyle C, Matthews R, Barlow J. Patient and public involvement: how much do we spend and what are the benefits? *Health Expect* 2015;**18**:1918–26. https://doi.org/10.1111/hex.12204

172. National Institute for Health and Care Excellence (NICE). *Referral Guidelines for Suspected Cancer. Clinical Guideline [CG27]*. London: NICE; 2005.

173. Lyratzopoulos G, Abel GA, McPhail S, Neal RD, Rubin GP. Measures of promptness of cancer diagnosis in primary care: secondary analysis of national audit data on patients with 18 common and rarer cancers. *Br J Cancer* 2013;**108**:686–90. https://doi.org/10.1038/bjc.2013.1

174. Whyte S, Harnan S, Scope A, Simpson E, Tappenden P, Duffy S, *et al. Early Awareness Interventions for Cancer: Colorectal Cancer*. URL: www.eepru.org.uk/article/early-awareness-interventions-for-cancer-colorectal-cancer-policy-research-unit-in-economic-evaluation-of-health-and-care-interventions/ (accessed 16 March 2019).

175. Murphy J, Halloran S, Gray A. Cost-effectiveness of the faecal immunochemical test at a range of positivity thresholds compared with the guaiac faecal occult blood test in the NHS Bowel Cancer Screening Programme in England. *BMJ Open* 2017;**7**:e017186. https://doi.org/10.1136/bmjopen-2017-017186

176. Office for National Statistics. *National Life Tables, UK: 2015 to 2017*. URL: www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/bulletins/nationallifetablesunitedkingdom/2015to2017 (accessed 16 March 2019).

177. Liu CY, Chen WT, Kung PT, Chiu CF, Wang YH, Shieh SH, Tsai WC. Characteristics, survival, and related factors of newly diagnosed colorectal cancer patients refusing cancer treatments under a universal health insurance program. *BMC Cancer* 2014;**14**:446. https://doi.org/10.1186/1471-2407-14-446

178. National Cancer Registration and Analysis Service. *Colorectal Cancer Survival by Stage – NCIN Data Briefing*. URL: www.ncin.org.uk/publications/data_briefings/colorectal_cancer_survival_by_stage (accessed 27 February 2019).

179. van Rijn JC, Reitsma JB, Stoker J, Bossuyt PM, van Deventer SJ, Dekker E. Polyp miss rate determined by tandem colonoscopy: a systematic review. *Am J Gastroenterol* 2006;**101**:343–50. https://doi.org/10.1111/j.1572-0241.2006.00390.x

180. Atkin WS, Cook CF, Cuzick J, Edwards R, Northover JM, Wardle J, UK Flexible Sigmoidoscopy Screening Trial Investigators. Single flexible sigmoidoscopy screening to prevent colorectal cancer: baseline findings of a UK multicentre randomised trial. *Lancet* 2002;**359**:1291–300. https://doi.org/10.1016/S0140-6736(02)08268-5

181. Gatto NM, Frucht H, Sundararajan V, Jacobson JS, Grann VR, Neugut AI. Risk of perforation after colonoscopy and sigmoidoscopy: a population-based study. *J Natl Cancer Inst* 2003;**95**:230–6. https://doi.org/10.1093/jnci/95.3.230

182. Curtis L, Burns A. *Unit Costs of Health and Social Care 2018*. Canterbury: PSSRU, University of Kent; 2018.

183. Ara R, Brazier JE. Populating an economic model with health state utility values: moving toward better practice. *Value Health* 2010;**13**:509–18. https://doi.org/10.1111/j.1524-4733.2010.00700.x

184. Ness RM, Holmes AM, Klein R, Dittus R. Utility valuations for outcome states of colorectal cancer. *Am J Gastroenterol* 1999;**94**:1650–7. https://doi.org/10.1111/j.1572-0241.1999.01157.x

185. Department of Health and Social Care (DHSC). *Improving Outcomes: A Strategy for Cancer*. London: DHSC; 2011.

186. Department of Health and Social Care (DHSC). *The Likely Impact of Earlier Diagnosis of Cancer on Costs and Benefits to the NHS*. London: DHSC; 2011. URL: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213788/dh_123576.pdf (accessed 16 March 2019).

187. Independent Cancer Taskforce. *Achieving World-Class Cancer Outcomes: Taking the Strategy Forward*. URL: www.england.nhs.uk/wp-content/uploads/2016/05/cancer-strategy.pdf (accessed 16 March 2019).

188. Vedsted P, Olesen F. Are the serious problems in cancer survival partly rooted in gatekeeper principles? An ecologic study. *Br J Gen Pract* 2011;**61**:e508–12. https://doi.org/10.3399/bjgp11X588484

189. Zhou Y, Mendonca SC, Abel GA, Hamilton W, Walter FM, Johnson S, *et al.* Variation in 'fast-track' referrals for suspected cancer by patient characteristic and cancer diagnosis: evidence from 670 000 patients with cancers of 35 different sites. *Br J Cancer* 2018;**118**:24–31. https://doi.org/10.1038/bjc.2017.381

190. NHS Digital. *GP Systems of Choice. Practice and System Data: Market Share 2015*. URL: http://content.digital.nhs.uk/media/20723/NIC-06225-H1B9Q—GPSoC-Market-Share-Practice-and-System-data—October-2015/xls/ (accessed 16 March 2019).

191. ClinRisk. *Welcome to QCancer*. URL: www.qcancer.org (accessed January 2020).

192. Cancer Research UK. *Cancer Risk Assessment Tools*. URL: www.cancerresearchuk.org/sites/default/files/rats.pdf (accessed January 2020).

193. Abel G, Saunders CL, Mendonca SC, Gildea C, McPhail S, Lyratzopoulos G. Variation and statistical reliability of publicly reported primary care diagnostic activity indicators for cancer: a cross-sectional ecological study of routine data. *BMJ Qual Saf* 2018;**27**:21–30. https://doi.org/10.1136/bmjqs-2017-006607

194. Kelley K, Clark B, Brown V, Sitzia J. Good practice in the conduct and reporting of survey research. *Int J Qual Health Care* 2003;**15**:261–6. https://doi.org/10.1093/intqhc/mzg031

195. Creavin ST, Creavin AL, Mallen CD. Do GPs respond to postal questionnaire surveys? A comprehensive review of primary care literature. *Fam Pract* 2011;**28**:461–7. https://doi.org/10.1093/fampra/cmr001

196. Abel GA, Barclay ME, Payne RA. Adjusted indices of multiple deprivation to enable comparisons within and between constituent countries of the UK including an illustration using mortality rates. *BMJ Open* 2016;**6**:e012750. https://doi.org/10.1136/bmjopen-2016-012750

197. Ministry of Housing, Communities & Local Government. *English Indices of Deprivation 2015: Technical Report*. URL: www.gov.uk/government/publications/english-indices-of-deprivation-2015-technical-report (accessed 16 March 2019).

198. Edwards PJ, Roberts I, Clarke MJ, Diguiseppi C, Wentz R, Kwan I, *et al*. Methods to increase response to postal and electronic questionnaires. *Cochrane Database Syst Rev* 2009;**3**:MR000008. https://doi.org/10.1002/14651858.MR000008.pub4

199. Hobbs FDR, Bankhead C, Mukhtar T, Stevens S, Perera-Salazar R, Holt T, *et al*. Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007–14. *Lancet* 2016;**387**:2323–30. https://doi.org/10.1016/S0140-6736(16)00620-6

200. Price SJ, Stapley SA, Shephard E, Barraclough K, Hamilton WT. Is omission of free text records a possible source of data loss and bias in Clinical Practice Research Datalink studies? A case–control study. *BMJ Open* 2016;**6**:e011664. https://doi.org/10.1136/bmjopen-2016-011664

201. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, *et al*. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;**21**:128–38. https://doi.org/10.1097/EDE.0b013e3181c30fb2

202. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;**26**:565–74. https://doi.org/10.1177/0272989X06295361

203. Debray TP, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KG. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *J Clin Epidemiol* 2015;**68**:279–89. https://doi.org/10.1016/j.jclinepi.2014.06.018

204. Adelstein BA, Irwig L, Macaskill P, Turner RM, Chan SF, Katelaris PH. Who needs colonoscopy to identify colorectal cancer? Bowel symptoms do not add substantially to age and other medical history. *Aliment Pharmacol Ther* 2010;**32**:270–81. https://doi.org/10.1111/j.1365-2036.2010.04344.x

205. Baicus C, Ionescu R, Tanasescu C. Does this patient have cancer? The assessment of age, anemia, and erythrocyte sedimentation rate in cancer as a cause of weight loss. A retrospective study based on a secondary care university hospital in Romania. *Eur J Intern Med* 2006;**17**:28–31. https://doi.org/10.1016/j.ejim.2005.07.009

206. Ewing M, Naredi P, Zhang C, Månsson J. Identification of patients with non-metastatic colorectal cancer in primary care: a case–control study. *Br J Gen Pract* 2016;**66**:e880–6. https://doi.org/10.3399/bjgp16X687985

207. Galvin R, Joyce D, Downey E, Boland F, Fahey T, Hill AK. Development and validation of a clinical prediction rule to identify suspected breast cancer: a prospective cohort study. *BMC Cancer* 2014;**14**:743. https://doi.org/10.1186/1471-2407-14-743

208. Khademi H, Radmard AR, 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**:e39173. https://doi.org/10.1371/journal.pone.0039173

209. Moore RG, Miller MC, Disilvestro P, Landrum LM, Gajewski W, Ball JJ, Skates SJ. Evaluation of the diagnostic accuracy of the risk of ovarian malignancy algorithm in women with a pelvic mass. *Obstet Gynecol* 2011;**118**:280–8. https://doi.org/10.1097/AOG.0b013e318224fce2

210. Saraiva R, Perkusich M, Silva L, Almeida H, Siebra C, Perkusich A. Early diagnosis of gastrointestinal cancer by using case-based and rule-based reasoning. *Expert Systems with Applications* 2016;**61**:192–202. https://doi.org/10.1016/j.eswa.2016.05.026

211. Reeves MJ, Osuch JR, Pathak DR. Development of a clinical decision rule for triage of women with palpable breast masses. *J Clin Epidemiol* 2003;**56**:636–45. https://doi.org/10.1016/S0895-4356(03)00123-9

212. Robertson R, Campbell C, Weller DP, Elton R, Mant D, Primrose J, *et al.* Predicting colorectal cancer risk in patients with rectal bleeding. *Br J Gen Pract* 2006;**56**:763–7.

213. Simpkins SJ, Pinto-Sanchez MI, Moayyedi P, Bercik P, Morgan DG, Bolino C, Ford AC. Poor predictive value of lower gastrointestinal alarm features in the diagnosis of colorectal cancer in 1981 patients in secondary care. *Aliment Pharmacol Ther* 2017;**45**:91–9. https://doi.org/10.1111/apt.13846

214. Koning NR, Moons LM, Büchner FL, Helsper CW, Ten Teije A, Numans ME. Identification of patients at risk for colorectal cancer in primary care: an explorative study with routine healthcare data. *Eur J Gastroenterol Hepatol* 2015;**27**:1443–8. https://doi.org/10.1097/MEG.0000000000000472

215. Kim MK, Kim K, Kim SM, Kim JW, Park NH, Song YS, Kang SB. A hospital-based case–control study of identifying ovarian cancer using symptom index. *J Gynecol Oncol* 2009;**20**:238–42. https://doi.org/10.3802/jgo.2009.20.4.238

216. Høgdall C, Fung ET, Christensen IJ, Nedergaard L, Engelholm SA, Petri AL, *et al.* A novel proteomic biomarker panel as a diagnostic tool for patients with ovarian cancer. *Gynecol Oncol* 2011;**123**:308–13. https://doi.org/10.1016/j.ygyno.2011.07.018

217. Dolianitis C, Kelly J, Wolfe R, Simpson P. Comparative performance of 4 dermoscopic algorithms by nonexperts for the diagnosis of melanocytic lesions. *Arch Dermatol* 2005;**141**:1008–14. https://doi.org/10.1001/archderm.141.8.1008

218. Argenziano G, Catricalà C, Ardigo M, Buccini P, De Simone P, Eibenschutz L, *et al.* Seven-point checklist of dermoscopy revisited. *Br J Dermatol* 2011;**164**:785–90. https://doi.org/10.1111/j.1365-2133.2010.10194.x

219. Emery JD, Hunter J, Hall PN, Watson AJ, Moncrieff M, Walter FM. Accuracy of SIAscopy for pigmented skin lesions encountered in primary care: development and validation of a new diagnostic algorithm. *BMC Dermatol* 2010;**10**:9. https://doi.org/10.1186/1471-5945-10-9

220. Gerbert B, Bronstone A, Maurer T, Hofmann R, Berger T. Decision support software to help primary care physicians triage skin cancer: a pilot study. *Arch Dermatol* 2000;**136**:187–92. https://doi.org/10.1001/archderm.136.2.187

221. Rogers T, Marino ML, Dusza SW, Bajaj S, Usatine RP, Marchetti MA, Marghoob AA. A clinical aid for detecting skin cancer: the Triage Amalgamated Dermoscopic Algorithm (TADA). *J Am Board Fam Med* 2016;**29**:694–701. https://doi.org/10.3122/jabfm.2016.06.160079

222. Walter FM, Morris HC, Humphrys E, Hall PN, Prevost AT, Burrows N, *et al.* Effect of adding a diagnostic aid to best practice to manage suspicious pigmented lesions in primary care: randomised controlled trial. *BMJ* 2012;**345**:e4110. https://doi.org/10.1136/bmj.e4110

223. Walter FM, Prevost AT, Vasconcelos J, Hall PN, Burrows NP, Morris HC, *et al.* Using the 7-point checklist as a diagnostic aid for pigmented skin lesions in general practice: a diagnostic validation study. *Br J Gen Pract* 2013;**63**:e345–53. https://doi.org/10.3399/bjgp13X667213

224. Ballal MS, Selvachandran SN, Maw A. Use of a patient consultation questionnaire and weighted numerical scoring system for the prediction of colorectal cancer and other colorectal pathology in symptomatic patients: a prospective cohort validation study of a Welsh population. *Colorectal Dis* 2010;**12**:407–14. https://doi.org/10.1111/j.1463-1318.2009.01984.x

225. Selvachandran SN, Hodder RJ, Ballal MS, Jones P, Cade D. Prediction of colorectal cancer by a patient consultation questionnaire and scoring system: a prospective study. *Lancet* 2002;**360**:278–83. https://doi.org/10.1016/S0140-6736(02)09549-1

226. Smith D, Ballal M, Hodder R, Soin G, Selvachandran SN, Cade D. Symptomatic presentation of early colorectal cancer. *Ann R Coll Surg Engl* 2006;**88**:185–90. https://doi.org/10.1308/003588406X94904

227. Zortea M, Schopf TR, Thon K, Geilhufe M, Hindberg K, Kirchesch H, *et al.* Performance of a dermoscopy-based computer vision system for the diagnosis of pigmented skin lesions compared with visual evaluation by experienced dermatologists. *Artif Intell Med* 2014;**60**:13–26. https://doi.org/10.1016/j.artmed.2013.11.006

228. Grewal K, Hamilton W, Sharp D. Ovarian cancer prediction: development of a scoring system for primary care. *BJOG* 2013;**120**:1016–19. https://doi.org/10.1111/1471-0528.12200

229. Shahzad N, Rashid N, Zahra S, Malik A. Role of malignancy index in prediction of malignancy in ovarian masses preoperative. *Med Forum Mon* 2015;**26**:44–7.

230. Rossing MA, Wicklund KG, Cushing-Haugen KL, Weiss NS. Predictive value of symptoms for early detection of ovarian cancer. *J Natl Cancer Inst* 2010;**102**:222–9. https://doi.org/10.1093/jnci/djp500

231. Lim AW, Mesher D, Gentry-Maharaj A, Balogun N, Jacobs I, Menon U, Sasieni P. Predictive value of symptoms for ovarian cancer: comparison of symptoms reported by questionnaire, interview, and general practitioner notes. *J Natl Cancer Inst* 2012;**104**:114–24. https://doi.org/10.1093/jnci/djr486

232. Law CW, Rampal S, Roslani AC, Mahadeva S. Development of a risk score to stratify symptomatic adults referred for colonoscopy. *J Gastroenterol Hepatol* 2014;**29**:1890–6. https://doi.org/10.1111/jgh.12638

233. Krown SE, Testa MA, Huang J. AIDS-related Kaposi's sarcoma: prospective validation of the AIDS Clinical Trials Group staging classification. AIDS Clinical Trials Group Oncology Committee. *J Clin Oncol* 1997;**15**:3085–92. https://doi.org/10.1200/JCO.1997.15.9.3085

234. Esteva M, Leiva A, Ramos M, Pita-Fernández S, González-Luján L, Casamitjana M, *et al.* Factors related with symptom duration until diagnosis and treatment of symptomatic colorectal cancer. *BMC Cancer* 2013;**13**:87. https://doi.org/10.1186/1471-2407-13-87

235. Smith EC, Ziogas A, Anton-Culver H. Delay in surgical treatment and survival after breast cancer diagnosis in young women by race/ethnicity. *JAMA Surg* 2013;**148**:516–23. https://doi.org/10.1001/jamasurg.2013.1680

236. Yun YH, Kim YA, Min YH, Park S, Won YJ, Kim DY, *et al.* The influence of hospital volume and surgical treatment delay on long-term survival after cancer surgery. *Ann Oncol* 2012;**23**:2731–7. https://doi.org/10.1093/annonc/mds101

237. Brazda A, Estroff J, Euhus D, Leitch AM, Huth J, Andrews V, *et al.* Delays in time to treatment and survival impact in breast cancer. *Ann Surg Oncol* 2010;**17**(Suppl. 3):291–6. https://doi.org/10.1245/s10434-010-1250-6

238. Eastman A, Tammaro Y, Moldrem A, Andrews V, Huth J, Euhus D, *et al.* Outcomes of delays in time to treatment in triple negative breast cancer. *Ann Surg Oncol* 2013;**20**:1880–5. https://doi.org/10.1245/s10434-012-2835-z

239. McLaughlin JM, Anderson RT, Ferketich AK, Seiber EE, Balkrishnan R, Paskett ED. Effect on survival of longer intervals between confirmed diagnosis and treatment initiation among low-income women with breast cancer. *J Clin Oncol* 2012;**30**:4493–500. https://doi.org/10.1200/JCO.2012.39.7695

240. Mujar M, Dahlui M, Yip CH, Tabib NA. Delays in time to primary treatment after a diagnosis of breast cancer: does it impact survival? *Prev Med* 2013;**56**:222–4. https://doi.org/10.1016/j.ypmed.2012.12.001

241. Redaniel MT, Martin RM, Cawthorn S, Wade J, Jeffreys M. The association of waiting times from diagnosis to surgery with survival in women with localised breast cancer in England. *Br J Cancer* 2013;**109**:42–9. https://doi.org/10.1038/bjc.2013.317

242. Sue GR, Lannin DR, Killelea B, Tsangaris T, Chagpar AB. Does time to definitive treatment matter in patients with ductal carcinoma in situ? *Am Surg* 2013;**79**:561–5.

243. Ermiah E, Abdalla F, Buhmeida A, Larbesh E, Pyrhönen S, Collan Y. Diagnosis delay in Libyan female breast cancer. *BMC Res Notes* 2012;**5**:452. https://doi.org/10.1186/1756-0500-5-452

244. Warner ET, Tamimi RM, Hughes ME, Ottesen RA, Wong YN, Edge SB, *et al.* Time to diagnosis and breast cancer stage by race/ethnicity. *Breast Cancer Res Treat* 2012;**136**:813–21. https://doi.org/10.1007/s10549-012-2304-1

245. Wright GP, Wong JH, Morgan JW, Roy-Chowdhury S, Kazanjian K, Lum SS. Time from diagnosis to surgical treatment of breast cancer: factors influencing delays in initiating treatment. *Am Surg* 2010;**76**:1119–22.

246. Wagner JL, Warneke CL, Mittendorf EA, Bedrosian I, Babiera GV, Kuerer HM, *et al.* Delays in primary surgical treatment are not associated with significant tumor size progression in breast cancer patients. *Ann Surg* 2011;**254**:119–24. https://doi.org/10.1097/SLA.0b013e318217e97f

247. Wallace DM, Bryan RT, Dunn JA, Begum G, Bathers S, West Midlands Urological Research Group. Delay and survival in bladder cancer. *BJU Int* 2002;**89**:868–78. https://doi.org/10.1046/j.1464-410X.2002.02776.x

248. Hollenbeck BK, Dunn RL, Ye Z, Hollingsworth JM, Skolarus TA, Kim SP, *et al.* Delays in diagnosis and bladder cancer mortality. *Cancer* 2010;**116**:5235–42. https://doi.org/10.1002/cncr.25310

249. Mommsen S, Aagaard J, Sell A. Presenting symptoms, treatment delay and survival in bladder cancer. *Scand J Urol Nephrol* 1983;**17**:163–7. https://doi.org/10.3109/00365598309180162

250. Gulliford MC, Petruckevitch A, Burney PG. Survival with bladder cancer, evaluation of delay in treatment, type of surgeon, and modality of treatment. *BMJ* 1991;**303**:437–40. https://doi.org/10.1136/bmj.303.6800.437

251. Tokuda Y, Chinen K, Obara H, Joishy SK. Intervals between symptom onset and clinical presentation in cancer patients. *Intern Med* 2009;**48**:899–905. https://doi.org/10.2169/internalmedicine.48.1720

252. Maguire A, Porta M, Malats N, Gallén M, Piñol JL, Fernandez E. Cancer survival and the duration of symptoms. An analysis of possible forms of the risk function. ISDS II Project Investigators. *Eur J Cancer* 1994;**30A**:785–92. https://doi.org/10.1016/0959-8049(94)90293-3

253. Liedberg F, Anderson H, Månsson A, Månsson W. Diagnostic delay and prognosis in invasive bladder cancer. *Scand J Urol Nephrol* 2003;**37**:396–400. https://doi.org/10.1080/00365590310006246

254. Thompson MR, Asiimwe A, Flashman K, Tsavellas G. Is earlier referral and investigation of bowel cancer patients presenting with rectal bleeding associated with better survival? *Colorectal Dis* 2011;**13**:1242–8. https://doi.org/10.1111/j.1463-1318.2010.02438.x

255. Terhaar sive Droste JS, Oort FA, van der Hulst RW, Coupé VM, Craanen ME, Meijer GA, *et al.* Does delay in diagnosing colorectal cancer in symptomatic patients affect tumor stage and survival? A population-based observational study. *BMC Cancer* 2010;**10**:332. https://doi.org/10.1186/1471-2407-10-332

256. Neal RD, Allgar VL, Ali N, Leese B, Heywood P, Proctor G, Evans J. Stage, survival and delays in lung, colorectal, prostate and ovarian cancer: comparison between diagnostic routes. *Br J Gen Pract* 2007;**57**:212–19.

257. Currie AC, Evans J, Smith NJ, Brown G, Abulafi AM, Swift RI. The impact of the two-week wait referral pathway on rectal cancer survival. *Colorectal Dis* 2012;**14**:848–53. https://doi.org/10.1111/j.1463-1318.2011.02829.x

258. Zafar A, Mak T, Whinnie S, Chapman MAS. The 2-week wait referral system does not improve 5-year colorectal cancer survival. *Colorectal Dis* 2012;**14**:e177–80. https://doi.org/10.1111/j.1463-1318.2011.02826.x

259. Gort M, Otter R, Plukker JT, Broekhuis M, Klazinga NS. Actionable indicators for short and long term outcomes in rectal cancer. *Eur J Cancer* 2010;**46**:1808–14. https://doi.org/10.1016/j.ejca.2010.02.044

260. Roland CL, Schwarz RE, Tong L, Ahn C, Balch GC, Yopp AC, *et al.* Is timing to delivery of treatment a reliable measure of quality of care for patients with colorectal adenocarcinoma? *Surgery* 2013;**154**:421–8. https://doi.org/10.1016/j.surg.2013.04.049

261. Van Hout AM, de Wit NJ, Rutten FH, Peeters PH. Determinants of patient's and doctor's delay in diagnosis and treatment of colorectal cancer. *Eur J Gastroenterol Hepatol* 2011;**23**:1056–63. https://doi.org/10.1097/MEG.0b013e32834c4839

262. Deng SX, An W, Gao J, Yin J, Cai QC, Yang M, *et al.* Factors influencing diagnosis of colorectal cancer: a hospital-based survey in China. *J Dig Dis* 2012;**13**:517–24. https://doi.org/10.1111/j.1751-2980.2012.00626.x

263. Cerdán-Santacruz C, Cano-Valderrama O, Cárdenas-Crespo S, Torres-García AJ, Cerdán-Miguel J. Colorectal cancer and its delayed diagnosis: have we improved in the past 25 years? *Rev Esp Enferm Dig* 2011;**103**:458–63. https://doi.org/10.4321/S1130-01082011000900004

264. Valentín-López B, Ferrándiz-Santos J, Blasco-Amaro JA, Morillas-Sáinz JD, Ruiz-López P. Assessment of a rapid referral pathway for suspected colorectal cancer in Madrid. *Fam Pract* 2012;**29**:182–8. https://doi.org/10.1093/fampra/cmr080

265. Ramsay G, MacKay C, Nanthakumaran S, Craig WL, McAdam TK, Loudon MA. Urgency of referral and its impact on outcome in patients with colorectal cancer. *Colorectal Dis* 2012;**14**:e375–7. https://doi.org/10.1111/j.1463-1318.2012.02961.x

266. Guzmán Laura KP, Bolíbar Ribas I, Alepuz MT, González D, Martín M. Impact on patient care time and tumor stage of a program for fast diagnostic and treatment of colorectal cancer. *Rev Esp Enferm Dig* 2011;**103**:13–19. https://doi.org/10.4321/s1130-01082011000100003

267. Tomlinson C, Wong C, Au HJ, Schiller D. Factors associated with delays to medical assessment and diagnosis for patients with colorectal cancer. *Can Fam Physician* 2012;**58**:e495–501.

268. Brocken P, Kiers BA, Looijen-Salamon MG, Dekhuijzen PN, Smits-van der Graaf C, Peters-Bax L, *et al.* Timeliness of lung cancer diagnosis and treatment in a rapid outpatient diagnostic program with combined 18FDG-PET and contrast enhanced CT scanning. *Lung Cancer* 2012;**75**:336–41. https://doi.org/10.1016/j.lungcan.2011.08.017

269. Radzikowska E, Roszkowski-Śliż K, Głaz P. The impact of timeliness of care on survival in non-small cell lung cancer patients. *Pneumonol Alergol Pol* 2012;**80**:422–9.

270. Loh LC, Chan LY, Tan RY, Govindaraju S, Ratnavelu K, Kumar S, *et al.* Time delay and its effect on survival in malaysian patients with non-small cell lung carcinoma. *Malays J Med Sci* 2006;**13**:37–42.

271. Annakkaya AN, Arbak P, Balbay O, Bilgin C, Erbas M, Bulut I. Effect of symptom-to-treatment interval on prognosis in lung cancer. *Tumori* 2007;**93**:61–7. https://doi.org/10.1177/030089160709300111

272. Skaug K, Eide GE, Gulsvik A. Predictors of long-term survival of lung cancer patients in a Norwegian community. *Clin Respir J* 2011;**5**:50–8. https://doi.org/10.1111/j.1752-699X.2010.00200.x

273. Pita-Fernández S, Montero-Martinez C, Pértega-Diaz S, Verea-Hernando H. Relationship between delayed diagnosis and the degree of invasion and survival in lung cancer. *J Clin Epidemiol* 2003;**56**:820–5. https://doi.org/10.1016/S0895-4356(03)00166-5

274. Gould MK, Ghaus SJ, Olsson JK, Schultz EM. Timeliness of care in veterans with non-small cell lung cancer. *Chest* 2008;**133**:1167–73. https://doi.org/10.1378/chest.07-2654

275. Diaconescu R, Lafond C, Whittom R. Treatment delays in non-small cell lung cancer and their prognostic implications. *J Thorac Oncol* 2011;**6**:1254–9. https://doi.org/10.1097/JTO.0b013e318217b623

276. González-Barcala FJ, García-Prim JM, Alvarez-Dobaño JM, Moldes-Rodríguez M, García-Sanz MT, Pose-Reino A, Valdés-Cuadrado L. Effect of delays on survival in patients with lung cancer. *Clin Transl Oncol* 2010;**12**:836–42. https://doi.org/10.1007/s12094-010-0606-5

277. Yilmaz A, Damadoglu E, Salturk C, Okur E, Tuncer LY, Halezeroglu S. Delays in the diagnosis and treatment of primary lung cancer: are longer delays associated with advanced pathological stage? *Ups J Med Sci* 2008;**113**:287–96. https://doi.org/10.3109/2000-1967-236

278. Christensen ED, Harvald T, Jendresen M, Aggestrup S, Petterson G. The impact of delayed diagnosis of lung cancer on the stage at the time of operation. *Eur J Cardiothorac Surg* 1997;**12**:880–4. https://doi.org/10.1016/S1010-7940(97)00275-3

279. Myrdal G, Lambe M, Hillerdal G, Lamberg K, Agustsson T, Ståhle E. Effect of delays on prognosis in patients with non-small cell lung cancer. *Thorax* 2004;**59**:45–9.

280. Murai T, Shibamoto Y, Baba F, Hashizume C, Mori Y, Ayakawa S, *et al.* Progression of non-small-cell lung cancer during the interval before stereotactic body radiotherapy. *Int J Radiat Oncol Biol Phys* 2012;**82**:463–7. https://doi.org/10.1016/j.ijrobp.2010.10.001

281. Salomaa ER, Sällinen S, Hiekkanen H, Liippo K. Delays in the diagnosis and treatment of lung cancer. *Chest* 2005;**128**:2282–8. https://doi.org/10.1378/chest.128.4.2282

282. Mohan A, Mohan C, Bhutani M, Pathak AK, Pal H, DAS C, Guleria R. Quality of life in newly diagnosed patients with lung cancer in a developing country: is it important? *Eur J Cancer Care (Engl)* 2006;**15**:293–8. https://doi.org/10.1111/j.1365-2354.2006.00654.x

283. Prabhu M, Kochupillai V, Sharma S, Ramachandran P, Sundaram KR, Bijlani L, Ahuja RK. Prognostic assessment of various parameters in chronic myeloid leukemia. *Cancer* 1986;**58**:1357–60. https://doi.org/10.1002/1097-0142(19860915)58:6<1357::aid-cncr2820580629>3.0.co;2-m

284. Friese CR, Earle CC, Magazu LS, Brown JR, Neville BA, Hevelone ND, *et al.* Timeliness and quality of diagnostic care for medicare recipients with chronic lymphocytic leukemia. *Cancer* 2011;**117**:1470–7. https://doi.org/10.1002/cncr.25655

285. Bertoli S, Bérard E, Huguet F, Huynh A, Tavitian S, Vergez F, *et al.* Time from diagnosis to intensive chemotherapy initiation does not adversely impact the outcome of patients with acute myeloid leukemia. *Blood* 2013;**121**:2618–26. https://doi.org/10.1182/blood-2012-09-454553

286. Jacobi N, Rogers TB, Peterson BA. Prognostic factors in follicular lymphoma: a single institution study. *Oncol Rep* 2008;**20**:185–93. https://doi.org/10.3892/or.20.1.185

287. Norum J. The effect of diagnostic delay in patients with Hodgkin's lymphoma. *Anticancer Res* 1995;**15**:2707–10.

288. Foulc P, N'Guyen JM, Dréno B. Prognostic factors in Sézary syndrome: a study of 28 patients. *Br J Dermatol* 2003;**149**:1152–8. https://doi.org/10.1111/j.1365-2133.2003.05677.x

289. Kim YH, Bishop K, Varghese A, Hoppe RT. Prognostic factors in erythrodermic mycosis fungoides and the Sézary syndrome. *Arch Dermatol* 1995;**131**:1003–8. https://doi.org/10.1001/archderm.1995.01690210033005

290. Kariyawasan CC, Hughes DA, Jayatillake MM, Mehta AB. Multiple myeloma: causes and consequences of delay in diagnosis. *QJM* 2007;**100**:635–40. https://doi.org/10.1093/qjmed/hcm077

291. Friese CR, Abel GA, Magazu LS, Neville BA, Richardson LC, Earle CC. Diagnostic delay and complications for older adults with multiple myeloma. *Leuk Lymphoma* 2009;**50**:392–400. https://doi.org/10.1080/10428190902741471

292. Koivunen P, Rantala N, Hyrynkangas K, Jokinen K, Alho OP. The impact of patient and professional diagnostic delays on survival in pharyngeal cancer. *Cancer* 2001;**92**:2885–91. https://doi.org/10.1002/1097-0142(20011201)92:11<2885::AID-CNCR10119>3.0.CO;2-G

293. Teppo H, Alho OP. Relative importance of diagnostic delays in different head and neck cancers. *Clin Otolaryngol* 2008;**33**:325–30. https://doi.org/10.1111/j.1749-4486.2008.01704.x

294. Teppo H, Koivunen P, Hyrynkangas K, Alho OP. Diagnostic delays in laryngeal carcinoma: professional diagnostic delay is a strong independent predictor of survival. *Head Neck* 2003;**25**:389–94. https://doi.org/10.1002/hed.10208

295. Wildt J, Bundgaard T, Bentzen SM. Delay in the diagnosis of oral squamous cell carcinoma. *Clin Otolaryngol Allied Sci* 1995;**20**:21–5. https://doi.org/10.1111/j.1365-2273.1995.tb00006.x

296. Seoane J, Takkouche B, Varela-Centelles P, Tomás I, Seoane-Romero JM. Impact of delay in diagnosis on survival to head and neck carcinomas: a systematic review with meta-analysis. *Clin Otolaryngol* 2012;**37**:99–106. https://doi.org/10.1111/j.1749-4486.2012.02464.x

297. Hansen O, Larsen S, Bastholt L, Godballe C, Jørgensen KE. Duration of symptoms: impact on outcome of radiotherapy in glottic cancer patients. *Int J Radiat Oncol Biol Phys* 2005;**61**:789–94. https://doi.org/10.1016/j.ijrobp.2004.07.720

298. McGurk M, Chan C, Jones J, O'regan E, Sherriff M. Delay in diagnosis and its effect on outcome in head and neck cancer. *Br J Oral Maxillofac Surg* 2005;**43**:281–4. https://doi.org/10.1016/j.bjoms.2004.01.016

299. Alho OP, Teppo H, Mäntyselkä P, Kantola S. Head and neck cancer in primary care: presenting symptoms and the effect of delayed diagnosis of cancer cases. *CMAJ* 2006;**174**:779–84. https://doi.org/10.1503/cmaj.050623

300. Sidler D, Thum P, Winterhalder R, Huber G, Haerle SK. Undifferentiated carcinoma of nasopharyngeal type (UCNT): a Swiss single-institutional experience during 1990-2005. *Swiss Med Wkly* 2010;**140**:273–9. https://doi.org/smw-12844

301. Caudell JJ, Locher JL, Bonner JA. Diagnosis-to-treatment interval and control of locoregionally advanced head and neck cancer. *Arch Otolaryngol Head Neck Surg* 2011;**137**:282–5. https://doi.org/10.1001/archoto.2011.20

302. Brouha XDR, Op De Coul B, Terhaard CH, Hordijk GJ. Does waiting time for radiotherapy affect local control of T1N0M0 glottic laryngeal carcinoma? *Clin Otolaryngol Allied Sci* 2000;**25**:215–18. https://doi.org/10.1046/j.1365-2273.2000.00347.x

303. Kumar S, Heller RF, Pandey U, Tewari V, Bala N, Oanh KT. Delay in presentation of oral cancer: a multifactor analytical study. *Natl Med J India* 2001;**14**:13–17.

304. Brouha XD, Tromp DM, Hordijk GJ, Winnubst JA, de Leeuw JR. Oral and pharyngeal cancer: analysis of patient delay at different tumor stages. *Head Neck* 2005;**27**:939–45. https://doi.org/10.1002/hed.20270

305. Sheng L, Shui Y, Shen L, Wei Q. Effect of patient-related delay in diagnosis on the extent of disease and prognosis in nasopharyngeal carcinoma. *Am J Rhinol* 2008;**22**:317–20. https://doi.org/10.2500/ajr.2008.22.3174

306. Tromp DM, Brouha XD, Hordijk GJ, Winnubst JA, de Leeuw RJ. Patient and tumour factors associated with advanced carcinomas of the head and neck. *Oral Oncol* 2005;**41**:313–19. https://doi.org/10.1016/j.oraloncology.2004.09.008

307. Allison P, Franco E, Black M, Feine J. The role of professional diagnostic delays in the prognosis of upper aerodigestive tract carcinoma. *Oral Oncol* 1998;**34**:147–53. https://doi.org/10.1016/S1368-8375(97)00088-2

308. Al-Rajhi N, El-Sebaie M, Khafaga Y, AlZahrani A, Mohamed G, Al-Amro A. Nasopharyngeal carcinoma in Saudi Arabia: clinical presentation and diagnostic delay. *East Mediterr Health J* 2009;**15**:1301–7.

309. Teppo H, Heikkinen J, Laitakari K, Alho OP. Diagnostic delays in vestibular schwannoma. *J Laryngol Otol* 2009;**123**:289–93. https://doi.org/10.1017/S0022215108003113

310. Pitchers M, Martin C. Delay in referral of oropharyngeal squamous cell carcinoma to secondary care correlates with a more advanced stage at presentation, and is associated with poorer survival. *Br J Cancer* 2006;**94**:955–8. https://doi.org/10.1038/sj.bjc.6603044

311. Vernham GA, Crowther JA. Head and neck carcinoma – stage at presentation. *Clin Otolaryngol Allied Sci* 1994;**19**:120–4. https://doi.org/10.1111/j.1365-2273.1994.tb01194.x

312. Lee AW, Foo W, Law SC, Poon YF, Sze WM, O SK, *et al.* Nasopharyngeal carcinoma: presenting symptoms and duration before diagnosis. *Hong Kong Med J* 1997;**3**:355–61.

313. Miziara ID, Cahali MB, Murakami MS, Figueiredo LA, Guimaraes JR. Cancer of the larynx: correlation of clinical characteristics, site of origin, stage, histology and diagnostic delay. *Rev Laryngol Otol Rhinol* 1998;**119**:101–4.

314. Scott SE, Grunfeld EA, McGurk M. The idiosyncratic relationship between diagnostic delay and stage of oral squamous cell carcinoma. *Oral Oncol* 2005;**41**:396–403. https://doi.org/10.1016/j.oraloncology.2004.10.010

315. Ho T, Zahurak M, Koch WM. Prognostic significance of presentation-to-diagnosis interval in patients with oropharyngeal carcinoma. *Arch Otolaryngol Head Neck Surg* 2004;**130**:45–51. https://doi.org/10.1001/archotol.130.1.45

316. Teppo H, Hyrynkangas K, Koivunen P, Jokinen K, Alho OP. Impact of patient and professional diagnostic delays on the risk of recurrence in laryngeal carcinoma. *Clin Otolaryngol* 2005;**30**:157–63. https://doi.org/10.1111/j.1365-2273.2004.00954.x

317. Umezu T, Shibata K, Kajiyama H, Yamamoto E, Mizuno M, Kikkawa F. Prognostic factors in stage IA-IIA cervical cancer patients treated surgically: does the waiting time to the operation affect survival? *Arch Gynecol Obstet* 2012;**285**:493–7. https://doi.org/10.1007/s00404-011-1966-y

318. Fruchter RG, Boyce J. Delays in diagnosis and stage of disease in gynecologic cancer. *Cancer Detect Prev* 1981;**4**:481–6.

319. Menczer J, Chetrit A, Sadetzki S, National Israel Ovarian Cancer Group. The effect of symptom duration in epithelial ovarian cancer on prognostic factors. *Arch Gynecol Obstet* 2009;**279**:797–801. https://doi.org/10.1007/s00404-008-0814-1

320. Crawford SC, Davis JA, Siddiqui NA, de Caestecker L, Gillis CR, Hole D, Penney G. The waiting time paradox: population based retrospective study of treatment delay and survival of women with endometrial cancer in Scotland. *BMJ* 2002;**325**:196. https://doi.org/10.1136/bmj.325.7357.196

321. Elit LM, O'Leary EM, Pond GR, Seow HY. Impact of wait times on survival for women with uterine cancer. *J Clin Oncol* 2014;**32**:27–33. https://doi.org/10.1200/JCO.2013.51.3671

322. Franceschi S, La Vecchia C, Gallus G, Decarli A, Colombo E, Mangioni C, Tognoni G. Delayed diagnosis of endometrial cancer in Italy. *Cancer* 1983;**51**:1176–8. https://doi.org/10.1002/1097-0142(19830315)51:6<1176::aid-cncr2820510634>3.0.co;2-o

323. Obermair A, Hanzal E, Schreiner-Frech I, Buxbaum P, Bancher-Todesca D, Thoma M, *et al.* Influence of delayed diagnosis on established prognostic factors in endometrial cancer. *Anticancer Res* 1996;**16**:947–9.

324. Pirog EC, Heller DS, Westhoff C. Endometrial adenocarcinoma – lack of correlation between treatment delay and tumor stage. *Gynecol Oncol* 1997;**67**:303–8. https://doi.org/10.1006/gyno.1997.4874

325. Robinson KM, Christensen KB, Ottesen B, Krasnik A. Diagnostic delay, quality of life and patient satisfaction among women diagnosed with endometrial or ovarian cancer: a nationwide Danish study. *Qual Life Res* 2012;**21**:1519–25. https://doi.org/10.1007/s11136-011-0077-3

326. Nagle CM, Francis JE, Nelson AE, Zorbas H, Luxford K, de Fazio A, *et al.* Reducing time to diagnosis does not improve outcomes for women with symptomatic ovarian cancer: a report from the Australian Ovarian Cancer Study Group. *J Clin Oncol* 2011;**29**:2253–8. https://doi.org/10.1200/JCO.2010.32.2164

327. Smith EM, Anderson B. The effects of symptoms and delay in seeking diagnosis on stage of disease at diagnosis among women with cancers of the ovary. *Cancer* 1985;**56**:2727–32. https://doi.org/10.1002/1097-0142(19851201)56:11<2727::aid-cncr2820561138>3.0.co;2-8

328. Lurie G, Wilkens LR, Thompson PJ, Matsuno RK, Carney ME, Goodman MT. Symptom presentation in invasive ovarian carcinoma by tumor histological type and grade in a multiethnic population: a case analysis. *Gynecol Oncol* 2010;**119**:278–84. https://doi.org/10.1016/j.ygyno.2010.05.028

329. Raptis DA, Fessas C, Belasyse-Smith P, Kurzawinski TR. Clinical presentation and waiting time targets do not affect prognosis in patients with pancreatic cancer. *Surgeon* 2010;**8**:239–46. https://doi.org/10.1016/j.surge.2010.03.001

330. McLean SR, Karsanji D, Wilson J, Dixon E, Sutherland FR, Pasieka J, *et al.* The effect of wait times on oncological outcomes from periampullary adenocarcinomas. *J Surg Oncol* 2013;**107**:853–8. https://doi.org/10.1002/jso.23338

331. Sundi D, Svatek RS, Margulis V, Wood CG, Matin SF, Dinney CP, Kamat AM. Upper tract urothelial carcinoma: impact of time to surgery. *Urol Oncol* 2012;**30**:266–72. https://doi.org/10.1016/j.urolonc.2010.04.002

332. Waldert M, Karakiewicz PI, Raman JD, Remzi M, Isbarn H, Lotan Y, *et al.* A delay in radical nephroureterectomy can lead to upstaging. *BJU Int* 2010;**105**:812–17. https://doi.org/10.1111/j.1464-410X.2009.08821.x

333. O'Brien D, Loeb S, Carvalhal GF, McGuire BB, Kan D, Hofer MD, *et al.* Delay of surgery in men with low risk prostate cancer. *J Urol* 2011;**185**:2143–7. https://doi.org/10.1016/j.juro.2011.02.009

334. Korets R, Seager CM, Pitman MS, Hruby GW, Benson MC, McKiernan JM. Effect of delaying surgery on radical prostatectomy outcomes: a contemporary analysis. *BJU Int* 2012;**110**:211–16. https://doi.org/10.1111/j.1464-410X.2011.10666.x

335. Sun M, Abdollah F, Hansen J, Trinh QD, Bianchi M, Tian Z, *et al.* Is a treatment delay in radical prostatectomy safe in individuals with low-risk prostate cancer? *J Sex Med* 2012;**9**:2961–9. https://doi.org/10.1111/j.1743-6109.2012.02806.x

336. Hanson PR, Belitsky P, Millard OH, Lannon SG. Prognostic factors in metastatic nonseminomatous germ cell tumours. *Can J Surg* 1993;**36**:537–40.

337. Fossa SD, Klepp O, Elgjo RF, Eliassen G, Melsom H, Urnes T, Wang M. The effect of patient's delay and doctor's delay in patients with malignant germ cell tumours. *Int J Androl* 1981;**4**(Suppl. 4):134–44. https://doi.org/10.1111/j.1365-2605.1981.tb00664.x

338. Huyghe E, Muller A, Mieusset R, Bujan L, Bachaud JM, Chevreau C, *et al.* Impact of diagnostic delay in testis cancer: results of a large population-based study. *Eur Urol* 2007;**52**:1710–16. https://doi.org/10.1016/j.eururo.2007.06.003

339. Moul JW, Paulson DF, Dodge RK, Walther PJ. Delay in diagnosis and survival in testicular cancer: impact of effective therapy and changes during 18 years. *J Urol* 1990;**143**:520–3. https://doi.org/10.1016/S0022-5347(17)40007-3

340. Harding M, Paul J, Kaye SB. Does delayed diagnosis or scrotal incision affect outcome for men with non-seminomatous germ cell tumours? *Br J Urol* 1995;**76**:491–4. https://doi.org/10.1111/j.1464-410x.1995.tb07754.x

341. Goh J, Bayati M, Zenios SA, Singh S, Moore D. Data uncertainty in Markov chains: application to cost-effectiveness analyses of medical innovations. *Oper Res* 2018;**66**:697–715. https://doi.org/10.1287/opre.2017.1685

342. Prout GR, Griffin PP. Testicular tumors: delay in diagnosis and influence on survival. *Am Fam Physician* 1984;**29**:205–9.

343. Dieckmann KP, Becker T, Bauer HW. Testicular tumors: presentation and role of diagnostic delay. *Urol Int* 1987;**42**:241–7. https://doi.org/10.1159/000281948

344. Meffan PJ, Delahunt B, Nacey JN. The value of early diagnosis in the treatment of patients with testicular cancer. *N Z Med J* 1991;**104**:393–4.

345. Bosl GJ, Vogelzang NJ, Goldman A, Fraley EE, Lange PH, Levitt SH, Kennedy BJ. Impact of delay in diagnosis on clinical stage of testicular cancer. *Lancet* 1981;**2**:970–3. https://doi.org/10.1016/S0140-6736(81)91165-X

346. Ware SM, Al-Askari S, Morales P. Testicular germ cell tumors. Prognostic factors. *Urology* 1980;**15**:348–52. https://doi.org/10.1016/0090-4295(80)90467-7

347. Wishnow KI, Johnson DE, Preston WL, Tenney DM, Brown BW. Prompt orchiectomy reduces morbidity and mortality from testicular carcinoma. *Br J Urol* 1990;**65**:629–33. https://doi.org/10.1111/j.1464-410x.1990.tb14834.x

348. Chilvers CE, Saunders M, Bliss JM, Nicholls J, Horwich A. Influence of delay in diagnosis on prognosis in testicular teratoma. *Br J Cancer* 1989;**59**:126–8. https://doi.org/10.1038/bjc.1989.25

349. Akdaş A, Kirkali Z, Remzi D. The role of delay in stage III testicular tumours. *Int Urol Nephrol* 1986;**18**:181–4. https://doi.org/10.1007/bf02082606

350. Scher H, Bosl G, Geller N, Cirrincione C, Whitmore W, Golbey R. Impact of symptomatic interval on prognosis of patients with stage III testicular cancer. *Urology* 1983;**21**:559–61. https://doi.org/10.1016/0090-4295(83)90190-5

351. Napier MP, Rustin GJ. Diagnostic delay and risk of relapse in patients with stage I nonseminomatous germ cell tumour followed on active surveillance. *BJU Int* 2000;**86**:486–9. https://doi.org/10.1046/j.1464-410X.2000.00779.x

352. Metzger S, Ellwanger U, Stroebel W, Schiebel U, Rassner G, Fierlbeck G. Extent and consequences of physician delay in the diagnosis of acral melanoma. *Melanoma Res* 1998;**8**:181–6. https://doi.org/10.1097/00008390-199804000-00014

353. Temoshok L, DiClemente RJ, Sweet DM, Blois MS, Sagebiel RW. Factors related to patient delay in seeking medical attention for cutaneous malignant melanoma. *Cancer* 1984;**54**:3048–53. https://doi.org/10.1002/1097-0142(19841215)54:12<3048::aid-cncr2820541239>3.0.co;2-m

354. Montella M, Crispo A, Grimaldi M, De Marco MR, Ascierto PA, Parasole R, *et al.* An assessment of factors related to tumor thickness and delay in diagnosis of melanoma in southern Italy. *Prev Med* 2002;**35**:271–7. https://doi.org/10.1006/pmed.2002.1067

355. Schmid-Wendtner MH, Baumert J, Stange J, Volkenandt M. Delay in the diagnosis of cutaneous melanoma: an analysis of 233 patients. *Melanoma Res* 2002;**12**:389–94. https://doi.org/10.1097/00008390-200208000-00012

356. Carli P, De Giorgi V, Palli D, Maurichi A, Mulas P, Orlandi C, *et al.* Dermatologist detection and skin self-examination are associated with thinner melanomas: results from a survey of the Italian Multidisciplinary Group on Melanoma. *Arch Dermatol* 2003;**139**:607–12. https://doi.org/10.1001/archderm.139.5.607

357. Baade PD, English DR, Youl PH, McPherson M, Elwood JM, Aitken JF. The relationship between melanoma thickness and time to diagnosis in a large population-based study. *Arch Dermatol* 2006;**142**:1422–7. https://doi.org/10.1001/archderm.142.11.1422

358. Helsing MD. Trofosfamide as a salvage treatment with low toxicity in malignant lymphoma. A phase II study. *Eur J Cancer* 1997;**33**:500–2. https://doi.org/10.1016/S0959-8049(97)89029-6

359. Richard MA, Grob JJ, Avril MF, Delaunay M, Thirion X, Wolkenstein P, *et al.* Melanoma and tumor thickness: challenges of early diagnosis. *Arch Dermatol* 1999;**135**:269–74. https://doi.org/10.1001/archderm.135.3.269

360. Krige JE, Isaacs S, Hudson DA, King HS, Strover RM, Johnson CA. Delay in the diagnosis of cutaneous malignant melanoma. A prospective study in 250 patients. *Cancer* 1991;**68**:2064–8. https://doi.org/10.1002/1097-0142(19911101)68:9<2064::aid-cncr2820680937>3.0.co;2-3

361. Cassileth BR, Clark WH, Heiberger RM, March V, Tenaglia A. Relationship between patients' early recognition of melanoma and depth of invasion. *Cancer* 1982;**49**:198–200. https://doi.org/10.1002/1097-0142(19820101)49:1<198::aid-cncr2820490138>3.0.co;2-9

362. Alam M, Goldberg LH, Silapunt S, Gardner ES, Strom SS, Rademaker AW, Margolis DJ. Delayed treatment and continued growth of nonmelanoma skin cancer. *J Am Acad Dermatol* 2011;**64**:839–48. https://doi.org/10.1016/j.jaad.2010.06.028

363. Renzi C, Mastroeni S, Passarelli F, Mannooranparampil TJ, Caggiati A, Potenza C, Pasquini P. Factors associated with large cutaneous squamous cell carcinomas. *J Am Acad Dermatol* 2010;**63**:404–11. https://doi.org/10.1016/j.jaad.2009.09.044

364. Lim BS, Dennis CR, Gardner B, Newman J. Analysis of survival versus patient and doctor delay of treatment in gastrointestinal cancer. *Am J Surg* 1974;**127**:210–14. https://doi.org/10.1016/0002-9610(74)90159-7

365. Ziliotto A, Kunzle JE, De Souza A, Colicchio-Filho O. Evolutive and prognostic aspects in gastric cancer. Analysis of 189 cases. *Cancer* 1987;**59**:811–17. https://doi.org/10.1002/1097-0142(19870215)59:4<811::aid-cncr2820590426>3.0.co;2-s

366. Arvanitakis C, Nikopoulos A, Giannoulis E, Theoharidis A, Georgilas V, Fotiou H, Tourkantonis A. The impact of early or late diagnosis on patient survival in gastric cancer in Greece. *Hepatogastroenterology* 1992;**39**:355–7.

367. Martin IG, Young S, Sue-Ling H, Johnston D. Delays in the diagnosis of oesophagogastric cancer: a consecutive case series. *BMJ* 1997;**314**:467–70. https://doi.org/10.1136/bmj.314.7079.467

368. Windham TC, Termuhlen PM, Ajani JA, Mansfield PF. Adenocarcinoma of the stomach in patients age 35 years and younger: no impact of early diagnosis on survival outcome. *J Surg Oncol* 2002;**81**:118–24. https://doi.org/10.1002/jso.10157

369. Fernandez E, Porta M, Malats N, Belloc J, Gallén M. Symptom-to-diagnosis interval and survival in cancers of the digestive tract. *Dig Dis Sci* 2002;**47**:2434–40. https://doi.org/10.1023/a:1020535304670

370. Maconi G, Kurihara H, Panizzo V, Russo A, Cristaldi M, Marrelli D, *et al.* Gastric cancer in young patients with no alarm symptoms: focus on delay in diagnosis, stage of neoplasm and survival. *Scand J Gastroenterol* 2003;**38**:1249–55. https://doi.org/10.1080/00365520310006360

371. Haugstvedt TK, Viste A, Eide GE, Søreide O. Patient and physician treatment delay in patients with stomach cancer in Norway: is it important? The Norwegian Stomach Cancer Trial. *Scand J Gastroenterol* 1991;**26**:611–19. https://doi.org/10.3109/00365529109043635

372. Wang J, Liu F, Gao H, Wei W, Zhang X, Liang Y, Cheng Y. The symptom-to-treatment delay and stage at the time of treatment in cancer of esophagus. *Jpn J Clin Oncol* 2008;**38**:87–91. https://doi.org/10.1093/jjco/hym169

373. Sharpe D, Williams RN, Ubhi SS, Sutton CD, Bowrey DJ. The 'two-week wait' referral pathway allows prompt treatment but does not improve outcome for patients with oesophago-gastric cancer. *Eur J Surg Oncol* 2010;**36**:977–81. https://doi.org/10.1016/j.ejso.2010.07.002

374. Grotenhuis BA, van Hagen P, Wijnhoven BP, Spaander MC, Tilanus HW, van Lanschot JJ. Delay in diagnostic workup and treatment of esophageal cancer. *J Gastrointest Surg* 2010;**14**:476–83. https://doi.org/10.1007/s11605-009-1109-y

375. Saithna A, Pynsent PB, Grimer RJ. Retrospective analysis of the impact of symptom duration on prognosis in soft tissue sarcoma. *Int Orthop* 2008;**32**:381–4. https://doi.org/10.1007/s00264-007-0319-8

376. Nakamura T, Matsumine A, Matsubara T, Asanuma K, Uchida A, Sudo A. The symptom-to-diagnosis delay in soft tissue sarcoma influence the overall survival and the development of distant metastasis. *J Surg Oncol* 2011;**104**:771–5. https://doi.org/10.1002/jso.22006

377. Rougraff BT, Davis K, Lawrence J. Does length of symptoms before diagnosis of sarcoma affect patient survival? *Clin Orthop Relat Res* 2007;**462**:181–9. https://doi.org/10.1097/BLO.0b013e3180f62608

378. Wurtz LD, Peabody TD, Simon MA. Delay in the diagnosis and treatment of primary bone sarcoma of the pelvis. *J Bone Joint Surg Am* 1999;**81**:317–25. https://doi.org/10.2106/00004623-199903000-00003

379. Ruka W, Emrich LJ, Driscoll DL, Karakousis CP. Tumor size/symptom duration ratio as a prognostic factor in patients with high-grade soft tissue sarcomas. *Eur J Cancer Clin Oncol* 1988;**24**:1583–8. https://doi.org/10.1016/0277-5379(88)90049-1

380. Bacci G, Ferrari S, Longhi A, Forni C, Zavatta M, Versari M, Smith K. High-grade osteosarcoma of the extremity: differences between localized and metastatic tumors at presentation. *J Pediatr Hematol Oncol* 2002;**24**:27–30. https://doi.org/10.1097/00043426-200201000-00008

381. Singal AG, Waljee AK, Patel N, Chen EY, Tiro JA, Marrero JA, Yopp AC. Therapeutic delays lead to worse survival among patients with hepatocellular carcinoma. *J Natl Compr Canc Netw* 2013;**11**:1101–8. https://doi.org/10.6004/jnccn.2013.0131

382. Holmäng S, Johansson SL. Impact of diagnostic and treatment delay on survival in patients with renal pelvic and ureteral cancer. *Scand J Urol Nephrol* 2006;**40**:479–84. https://doi.org/10.1080/00365590600864093

383. Toth-Fejel S, Pommier RF. Relationships among delay of diagnosis, extent of disease, and survival in patients with abdominal carcinoid tumors. *Am J Surg* 2004;**187**:575–9. https://doi.org/10.1016/j.amjsurg.2004.01.019

384. Edge SB, Compton CC. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann Surg Oncol* 2010;**17**:1471–4. https://doi.org/10.1245/s10434-010-0985-4

385. Zinkin LD. A critical review of the classifications and staging of colorectal cancer. *Dis Colon Rectum* 1983;**26**:37–43. https://doi.org/10.1007/bf02554677

386. Bolin TD, Korman MG, Stanton R, Talley NJ, Newstead GL, Donnelly N, *et al.* Positive cost effectiveness of early diagnosis of colorectal cancer. *Colorectal Dis* 1999;**1**:113–22. https://doi.org/10.1046/j.1463-1318.1999.00028.x

387. Byers T, Gorsky R. Estimates of costs and effects of screening for colorectal cancer in the United States. *Cancer* 1992;**70**(Suppl. 5):1288–95. https://doi.org/10.1002/1097-0142(19920901)70:3+<1288::aid-cncr2820701515>3.0.co;2-1

388. Neilson AR, Whynes DK. Cost-effectiveness of screening for colorectal cancer: a simulation model. *IMA J Math Appl Med Biol* 1995;**12**:355–67. https://doi.org/10.1093/imammb/12.3-4.355

389. Gow J. Costs of screening for colorectal cancer: an Australian programme. *Health Econ* 1999;**8**:531–40. https://doi.org/10.1002/(SICI)1099-1050(199909)8:6<531::AID-HEC441>3.0.CO;2-K

390. Khandker RK, Dulski JD, Kilpatrick JB, Ellis RP, Mitchell JB, Baine WB. A decision model and cost-effectiveness analysis of colorectal cancer screening and surveillance guidelines for average-risk adults. *Int J Technol Assess Health Care* 2000;**16**:799–810. https://doi.org/10.1017/s0266462300102077

391. Sonnenberg A, Delcò F, Inadomi JM. Cost-effectiveness of colonoscopy in screening for colorectal cancer. *Ann Intern Med* 2000;**133**:573–84. https://doi.org/10.7326/0003-4819-133-8-200010170-00007

392. Reyes CM, Allen BA, Terdiman JP, Wilson LS. Comparison of selection strategies for genetic testing of patients with hereditary nonpolyposis colorectal carcinoma: effectiveness and cost-effectiveness. *Cancer* 2002;**95**:1848–56. https://doi.org/10.1002/cncr.10910

393. Kievit W, de Bruin JH, Adang EM, Severens JL, Kleibeuker JH, Sijmons RH, *et al.* Cost effectiveness of a new strategy to identify HNPCC patients. *Gut* 2005;**54**:97–102. https://doi.org/10.1136/gut.2004.039123

394. Telford JJ, Levy AR, Sambrook JC, Zou D, Enns RA. The cost-effectiveness of screening for colorectal cancer. *CMAJ* 2010;**182**:1307–13. https://doi.org/10.1503/cmaj.090845

395. Allameh Z, Davari M, Emami MH. Cost-effectiveness analysis of colorectal cancer screening methods in Iran. *Arch Iran Med* 2011;**14**:110–14. https://doi.org/011142/AIM.008

396. Dan YY, Chuah BY, Koh DC, Yeoh KG. Screening based on risk for colorectal cancer is the most cost-effective approach. *Clin Gastroenterol Hepatol* 2012;**10**:266–71.e1–6. https://doi.org/10.1016/j.cgh.2011.11.011

397. Atkin W, Brenner A, Martin J, Wooldrage K, Shah U, Lucas F, *et al.* The clinical effectiveness of different surveillance strategies to prevent colorectal cancer in people with intermediate-grade colorectal adenomas: a retrospective cohort analysis, and psychological and economic evaluations. *Health Technol Assess* 2017;**21**(25). https://doi.org/10.3310/hta21250

398. Office for National Statistics. *Cancer Diagnoses and Age-standardised Incidence Rates for all Types of Cancer by Age, Sex and Region Including Breast, Prostate, Lung and Colorectal Cancer*. URL: www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancerregistrationstatisticsengland/2017 (accessed 1 March 2019).

399. NHS Improvement. *Archives Reference Costs*. 2020. URL: https://improvement.nhs.uk/resources/reference-costs/ (accessed December 2020).

400. GOV.UK. *NHS Reference Costs 2014 to 2015*. 2015. URL: www.gov.uk/government/publications/nhs-reference-costs-2014-to-2015 (accessed December 2020).

401. Gavin DR, Valori RM, Anderson JT, Donnelly MT, Williams JG, Swarbrick ET. The national colonoscopy audit: a nationwide assessment of the quality and safety of colonoscopy in the UK. *Gut* 2013;**62**:242–9. https://doi.org/10.1136/gutjnl-2011-301848

402. Ara R, Brazier J. Estimating health state utility values for comorbid health conditions using SF-6D data. *Value Health* 2011;**14**:740–5. https://doi.org/10.1016/j.jval.2010.12.011

403. National Institute for Health and Care Excellence (NICE). *Colorectal Cancer*. London: NICE; 2020

404. Whyte S, Chilcott J, Essat M, Stevens J, Wong R, Kalita N. *Re-appraisal of the Options for Colorectal Cancer Screening. Report for the NHS Bowel Cancer Screening Programme*. Sheffield: School of Health and Related Research, University of Sheffield; 2011.

Appendix 1 Literature search strategies

Database

MEDLINE

Host: Ovid.

Date range searched: 1946 to May week 1 2017.

Date searched: 11 May 2017.

Searcher: SR.

Hits: 2008.

Search strategy

1. exp Neoplasms/
2. (cancer$ or neopla$).tw.
3. (tumour$ or tumor$).tw.
4. or/1-3
5. Primary Health Care/
6. exp General Practice/
7. General Practitioners/
8. (primary care or general practi$ or family practi$).tw.
9. (primary adj3 (healthcare or health care)).tw.
10. or/5-9
11. Decision Support Systems, Clinical/
12. Decision Support Techniques/
13. (tool or tools or aid$ or model or models or checklist$ or check list$ or rule or rules or algorithm$ or ation$).tw.
14. or/11-13
15. "Early Detection of Cancer"/
16. (predict$ or assess$ or scor$ or risk$ or validat$ or decision$ or identif$ or diagno$ or prognos$).tw.
17. (2ww or 2 week wait or two week wait or 2 week rule or two week rule).tw.
18. or/15-17
19. 4 and 10 and 14 and 18.

MEDLINE In-Process & Other Non-Indexed Citations

Host: Ovid

Date range searched: 10 May 2017.

Date searched: 11 May 2017.

Searcher: SR.

Hits: 270.

Search strategy

1. (cancer$ or neopla$).tw.
2. (tumour$ or tumor$).tw.
3. or/1-2
4. primary care or general practi$ or family practi$).tw.
5. (primary adj3 (healthcare or health care)).tw.
6. or/4-5
7. (tool or tools or aid$ or model or models or checklist$ or check list$ or rule or rules or algorithm$ or equation$).tw.
8. (predict$ or assess$ or scor$ or risk$ or validat$ or decision$ or identif$ or diagno$ or prognos$).tw.
9. (2ww or 2 week wait or two week wait or 2 week rule or two week rule).tw.
10. or/8-9
11. 3 and 6 and 7 and 10.

EMBASE

Host: Ovid.

Date range searched: 1974 to 10 May 2017.

Date searched: 11 May 2017.

Searcher: SR.

Hits: 3867.

Search strategy

1. exp Neoplasms/
2. (cancer$ or neopla$).tw.
3. (tumour$ or tumor$).tw.
4. or/1-3
5. Primary Health Care/
6. exp General Practice/
7. General Practitioners/
8. (primary care or general practi$ or family practi$).tw.
9. (primary adj3 (healthcare or health care)).tw.
10. or/5-9
11. (tool or tools or aid$ or model or models or checklist$ or check list$ or rule or rules or algorithm$ or equation$).tw.
12. exp decision support system/
13. 11 or 12
14. (predict$ or assess$ or scor$ or risk$ or validat$ or decision$ or identif$ or diagnos$ or prognos$).tw.
15. (2ww or 2 week wait or two week wait or 2 week rule or two week rule).tw.
16. Early Cancer Diagnosis/
17. or/14-16
18. 4 and 10 and 13 and 17.

Cochrane Database of Systematic Reviews and Cochrane Central Register of Controlled Trials

Host: Cochrane Library.

Date ranges searched: Cochrane Database of Systematic Reviews (CDSR) – issue 5 of 12, May 2017; Cochrane Central Register of Controlled Trials – issue 4 of 12, April 2017.

Date searched: 11 May 2017.

Searcher: SR.

Hits: 391.

Search strategy

1. MeSH descriptor: [Neoplasms] explode all trees
2. (cancer* or neopla*):ti,ab,kw
3. (tumour* or tumor*):ti,ab,kw
4. #1 or #2 or #3
5. MeSH descriptor: [Primary Health Care] explode all trees
6. MeSH descriptor: [General Practice] explode all trees
7. MeSH descriptor: [General Practitioners] explode all trees
8. ("primary care" or "general practi*" or "family practi*"):ti,ab,kw
9. (primary near/3 (healthcare or "health care")):ti,ab,kw
10. #5 or #6 or #7 or #8 or #9
11. MeSH descriptor: [Decision Support Techniques] this term only
12. MeSH descriptor: [Decision Support Systems, Clinical] this term only
13. (tool or tools or aid* or model or models or checklist* or "check list" * or rule or rules or algorithm* or equation*):ti,ab,kw
14. #11 or #12 or #13
15. MeSH descriptor: [Early Detection of Cancer] this term only
16. (predict* or assess* or scor* or risk* or validat* or decision* or identif* or diagnos* or prognos*):ti,ab,kw
17. (2ww or "2 week wait" or "two week wait" or "2 week rule" or "two week rule"):ti,ab,kw
18. #15 or #16 or #17
19. #4 and #10 and #14 and #18.

Science Citation Index and Conference Proceedings Citation Index – Science

Host: Web of Science.

Date range searched: not applicable.

Date searched: 11 May 2017.

Searcher: SR.

Hits: 2816.

Search strategy

1. TS=((cancer* or neopla*))
2. TS=((tumour* or tumor*))
3. #1 or #2
4. TS=(("primary care" or "general practi*" or "family practi*"))
5. TS=((primary near/2 healthcare))
6. TS=((primary near/2 "health care"))
7. #3 or #4 or #5

8. TS= ((tool or tools or aid* or model or models or checklist* or "check list*" or rule or rules or algorithm* or equation*))
9. TS=((predict* or assess* or scor* or risk* or validat* or decision* or identif* or diagnos* or prognos*))
10. TS=((2ww or "2 week wait" or "two week wait" or "2 week rule" or "two week rule"))
11. #9 or #10
12. #3 and #7 and #8 and #11
13. Indexes=SCI-EXPANDED, CPCI-S Timespan=2004-2011.

DocType=All document types; Language=All languages.

Search results

The following table shows the number of hits per database and in total.

Database	Hits (*n*)
MEDLINE	2008
MEDLINE In-Process & Other Non-Indexed Citations	270
EMBASE	3867
Cochrane Library (CDSR and CENTRAL)	391
Web of Science (SCI and CPCI-S)	2816
Total records	9352
Duplicates	3743
Total unique records	5609

SCI, Science Citation Index.

The updated searches were carried out on 30 January 2018. The next table shows the number of hits per database and in total for the updated searches.

Database	Hits (*n*)
MEDLINE	168
MEDLINE In-Process & Other Non-Indexed Citations	1
EMBASE	336
Cochrane Library (CDSR and CENTRAL)	29
Web of Science (SCI and CPCI-S)	181
Total records	715
Duplicates	84
Total unique records	631

SCI, Science Citation Index.

Appendix 2 Systematic review 1: supplementary tables

TABLE 33 Systematic review 1: included studies, study characteristics

Study	Prediction tool	Cancer type(s)	Country	Study design	Intended purpose
Del Mar 1995[37]	Melanoma 'algorithm' (plus camera)	Melanoma	Australia	Field trial	To assess the algorithm's ability to reduce the number of benign lesions being excised without reducing the excision of invasive lesions by comparing the numbers of excised lesions with algorithm use with the number excised without algorithm use
English 2003[36]	Melanoma 'algorithm' (plus camera)	Melanoma	Australia	RCT	To determine whether or not an aid to the diagnosis of pigmented skin lesions reduces the ratio of benign lesions to melanomas excised after introduction in general practice
Gulati 2015[38]	GP Skin Cancer Toolkit	Skin cancer	UK	Case–control	To assess the impact of the toolkit by comparing before and after national skin cancer referral data, cross-sectional questionnaires and urgent skin cancer referral data to two NHS trusts
Hamilton 2013[13]	RAT for lung cancer and CRC in two formats: mouse mat and desktop flip chart	Lung cancer, CRC	UK	Cohort study	To compare referrals and investigations for colorectal and lung cancer before and after the implementation of a RAT
Emery 2017[35]	Composite intervention including the RAT for colorectal, lung and prostate cancers, as well as summaries of relevant guidelines for colorectal, lung, prostate and breast cancers	CRC, lung cancer, prostate cancer, breast cancer	Australia	Factorial cluster RCT	To measure the effect of community-based symptom awareness and general practice-based educational interventions on the time to diagnosis in rural patients presenting with breast, prostate, colorectal or lung cancer in Western Australia

TABLE 34 Systematic review 1: included studies, study design

Study	Population	Recruitment (plus inclusion and exclusion criteria)	Sample size	Tool description	Predictors
Del Mar 1995[37]	Medical practitioners, mostly in general practice, in each of two cities in tropical Queensland, Australia Examination of the histopathological reports of 5823 melanocytic skin lesions excised over the intervention period (2 years) and in the preceding 6 months	Two provincial cities in central Queensland were selected for the trial on the basis of their similarity; one city was randomly selected to implement the use of algorithm, the other city was used as a control	53 medical practitioners (45 GPs, seven surgeons and one dermatologist) in the control community; 52 medical practitioners (48 GPs and four surgeons) in the intervention community	An algorithm for managing clinically suspicious naevi, aided by the use of a camera	Change in lesion characteristics: size, outline, colour, elevation, surface change, daughter lesions, tingling or itching; aged > 40 years
English 2003[36]	Data on examined and/or excised skin lesions suspected of being malignant, reported by general practices in Perth, WA, Australia	General practices from Perth, WA, Australia, were randomised to either control (national guidelines on managing melanoma) or intervention (diagnosis tool)	• 223 randomised practices (468 GPs) • Intervention: 111 practices (245 GPs) • Control: 112 practices (228 GPs)	An algorithm for managing clinically suspicious naevi, aided by the use of a camera	Change in lesion characteristics: size, outline, colour, elevation, surface change, daughter lesions, tingling or itching; and aged > 40 years
Gulati 2015[38]	2WW referral data from 2456 practices were obtained for 4 months during the launch of the toolkit (July–October 2012) and compared with referral habits in the same months in the previous year (July–October 2011). Referral habits during these two time periods were also measured in 3693 practices where the GP Skin Cancer Toolkit was not used. Appropriateness of the 2WW referrals based on data from Homerton University Hospital and Royal London Hospital between July and October 2011 and between July and October 2012	Data on participating practices were based on website usage data (www.doctors.net.uk). Cancer Waiting Times data were obtained through PHE's National Cancer Registration Service (East Midlands) and taken from the National Cancer Waiting Times Monitoring Dataset, provided by NHS England. A total of 276 GPs referring to Barts Health NHS Trust and Homerton University Hospital NHS Foundation Trust were analysed for appropriateness according to NICE referral guidelines	Impact on referral statistics: 2456 practices that used the toolkit as cases; 3693 practices as controls. Appropriateness of urgent referral: 2011 – 365 referrals from 214 GPs; 2012 – 387 referrals from 183 GPs (114 GPs had access to toolkit)	The toolkit consisted of a referral decision aid (referral guidelines based on red flags), lesion recognition resource (a series of images), clinical cases and a quiz	NICE referral guidelines (weighted seven-point checklist)

Study	Population	Recruitment (plus inclusion and exclusion criteria)	Sample size	Tool description	Predictors
Hamilton 2013[13]	Number of investigations (X-rays or colonoscopies) and 2WW referrals from practices and local NHS trusts for the two 6-month periods before and after the distribution of the tools	A selected GP cancer lead from seven of the 28 English cancer networks recruited local general practices to which the RATs were supplied. A total of 614 GPs from 165 practices were recruited; 2593 assessments were included	614 GPs from 165 practices	RAT gives risk estimates for patients aged > 40 years presenting to primary care with symptoms of possible cancer, for single symptoms, pairs of symptoms and repeat attendances with the same symptom. The values are colour coded to aid interpretation	For CRC: symptoms (constipation, diarrhoea, rectal bleeding, abdominal pain, abdominal tenderness, abnormal rectal examination), loss of weight, haemoglobin (levels 10–13 g/dl; < 10 g/dl); NR for other cancers
Emery 2017[35]	Two trial areas in Western Australia	Inclusion criteria: adults aged > 18 years; diagnosed with breast, lung, colorectal or prostate cancer between 1 January 2012 and the recruitment end date of 31 March 2014; and resident of trial areas at the time of cancer diagnosis	1358 patients (497 in trial area A, 861 in trial area B)	Resource card containing the RAT tables for colorectal, lung and prostate cancers, as well as the National Breast and Ovarian Cancer Centre guidelines for investigating new breast symptoms	NR

NR, not reported; PHE, Public Health England.

TABLE 35 Systematic review 1: included studies, results

Study	Outcomes (definitions)	Outcomes	Main results
Del Mar 1995[37]	Percentages of benign (neither malignant nor potentially malignant) melanocytic lesions excised during the 2-year intervention period	Number of excisions performed in each city was measured at baseline (6 months before implementation of the intervention) and 2 years after; percentage of excised lesions that are neither invasive nor potentially malignant; and median number of excisions per doctor	Total number of excised lesions: • At baseline – control city, 752; intervention city, 606 • After intervention – control city, 2468; intervention city, 1997 Percentage of excised lesions that are neither invasive nor potentially malignant: • At baseline – control city, 94% (95% CI 92.3% to 95.7%); intervention city, 93.6% (95% CI 91.6% to 95.5%) ($p = 0.731$, χ^2 test) • After intervention – control city, 93.8% (95% CI 92.8% to 94.8%); intervention city, 88.8% (95% CI 87.4% to 90.2%) The median number of excisions per doctor (2.5% and 97.5% percentiles) • At baseline – control city, 7 (1, 38); intervention city, 8 (1, 46) • After intervention – control city, 9 (1, 39); intervention city, 4 (1, 30)

English 2003[36]	Malignant lesion – an in situ or invasive melanoma benign lesion – naevus (including dysplastic naevus) or a seborrhoeic keratosis	Ratio of benign pigmented lesions to melanomas excised	Number of excised skin lesions (including seborrheic keratoses):
			• At baseline –
			  ○ Control: benign, 1965; melanoma, 61; ratio 32
			  ○ Intervention: benign, 2615; melanoma, 100; ratio 26
			• Trial period –
			  ○ Control: benign, 2037; melanoma, 79
			  ○ Intervention: benign, 2369; melanoma, 81; 1.03 times higher (95% CI 0.71 to 1.50, $p = 0.88$)

Gulati 2015[38]	Number of urgent (2WW) cancer referrals; proportion of appropriate urgent referrals (based on the NICE guidelines); melanoma and non-melanoma skin cancer rates	Number of urgent referrals, diagnoses of melanoma skin cancer; appropriateness of urgent referrals	21,000 GPs were invited to use the tool; 8163 GPs accessed the tool during the 2012 period. There were no significant changes in the number of urgent GP referrals for suspected skin cancer (Spearman's rank 0.20; $p < 0.001$), diagnoses of melanoma (Spearman's rank 0.064; $p < 0.001$) or diagnoses of non-melanoma skin cancer (Spearman's rank 0.068; $p < 0.001$) between the toolkit user and non-user groups. The proportion of appropriate referrals increased from 21.37% in 2011 to 32.3% in 2012, giving an incidence rate ratio of 3.13 (95% CI 2.21 to 4.42, z-statistic 6.46; $p < 0.0001$); the differences in numbers of appropriate referrals between toolkit users and non-users did not reach statistical significance by Spearman's rank test and ANOVA

TABLE 35 Systematic review 1: included studies, results (*continued*)

Study	Outcomes (definitions)	Outcomes	Main results
Hamilton 2013[13]	Investigation for cancer: 2-week referral to the appropriate specialty, or a chest X-ray in possible lung cancer	Number of investigations and two-week referrals	Lung cancer: 31% increase in 2-week referrals (332 before, 436 after); 4% increase in related investigations (chest X-ray: 7431 before, 7723 after) Colorectal cancer: 26% increase in 2-week referrals (1173 before, 1477 after); 15% increase in colonoscopies (1762 before, 2032 after)
Emery 2017[35]	TDI, as the time from first symptom to cancer diagnosis	TDI, as the time from first symptom to cancer diagnosis	No significant differences in the median or ln mean TDI at either intervention level: • GP intervention vs. control – median TDI 97 vs. 96.5 days; ln mean difference 0.004 (95% CI 0.18 to 0.19; $p = 0.99$) • Community intervention vs. control – median TDI 107.5 vs. 92 days; ln mean difference 0.08 (95% CI 0.06 to 0.23; $p = 0.27$) No significant differences in the TDI when analysed by factorial design, tumour group or subintervals of the TDI

ANOVA, analysis of variance.

TABLE 36 Systematic review 1: included studies, critical appraisal

Study	Risk of bias								
	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
Del Mar 1995[37]	High risk (selection of intervention site was arbitrary)	Low risk (sites chosen sufficiently far apart that an intervention in one site was unlikely to affect clinical behaviour in the other)	Low risk	Low risk	High risk (a large proportion of GPs did not participate post randomisation)	High risk (use of the intervention was not blinded)	Low risk ('data analysis was performed before the code identifying the city was broken')	Low risk	Unclear risk
English 2003[36]	Low risk (intervention practices were randomly selected)	Low risk (randomisation codes in sealed envelopes)	Low risk	Low risk	Unclear risk	High risk ('after randomisation, participants and research assistants who visited practices were not blinded to assignment')	Low risk ('all coding of outcome data was done blind to assignment')	Low risk	Unclear risk
Gulati 2015[38]	High risk (voluntary selection of participants)	High risk (use of the intervention was not blinded)	Low risk	Low risk	Unclear risk	Unclear risk (based on reported use of the toolkit)	Unclear risk	Unclear risk	High risk (only 39% of GPs who had access to the toolkit used it)
Hamilton 2013[13]	High risk (not a RCT)	High risk (allocation not concealed)	N/A (cohort study)	N/A (cohort study)	Unclear risk (details not reported)	High risk (the GPs were not blinded, irrespective of whether or not they were counted)	Unclear risk (the intervention was not blinded)	Unclear risk	High risk (other important influencing factors could have occurred during the before and after. Unclear how many GPs involved)
Emery 2017[35]	Low risk (allocation of interventions was random)	High risk (allocation not concealed)	Low risk	Low risk	Low risk	Low risk (research staff who collected outcome data and the trial statistician were blinded to group allocation)	Low risk (measures were taken to avoid contamination: re-clustering of practices attended by same practitioners, use of media avoided in the control areas)	Low risk	Low risk

N/A, not applicable.

Appendix 3 Systematic review 2: supplementary tables

TABLE 37 Systematic review 2: included studies, study characteristics (1)

Study	Prediction model	Cancer type(s)	Country	Setting	Study design	Stage of development	Data source
Bladder							
Shephard 2012[69]	RAT	Bladder	UK	Primary care	Case–control	Apparent performance	GPRD (now called the CPRD)
Blood							
Shephard 2016[67]	RAT	Leukaemia	UK	Primary care	Case–control	Apparent performance	GPRD (now called the CPRD)
Shephard 2015[68]	RAT	Multiple myeloma	UK	Primary care	Case–control	Apparent performance	GPRD (now called the CPRD)
Brain							
Hamilton 2007[87]	RAT	Brain	UK	Primary care	Case–control	Apparent performance	GPRD (now called the CPRD)
Breast							
McCowan 2011[60]	Clinical prediction rule for breast cancer	Breast	UK	Developed in secondary care, for use in primary care	Prospective cohort	External validation	Derivation cohort: consecutive patients attending the symptomatic breast clinic at Ninewells Hospital, Dundee (Scotland), who were referred by a GP between February and June 2007 Validation cohort: all women who attended for an initial consultation regarding symptomatic breast problems in 11 participating general practices in the region between January 2006 and June 2007
Colorectal							
Marshall 2011[22]	BB equation	Colorectal	UK	Primary care	Case–control	External validation	Derivation cohort: THIN Validation cohort: CAPER
Elias 2017[50]	BB equation	Colorectal	The Netherlands	Primary care	Prospective cross-sectional	External validation	CEDAR study: patients referred to endoscopy centres by participating Dutch primary care practices. 2009–12

Study	Prediction model	Cancer type(s)	Country	Setting	Study design	Stage of development	Data source
Fijten 1995[55]	Referred to as the 'Netherlands model' in Hodder 2005[77] and Elias 2017[50]	Colorectal	The Netherlands	Primary care	Prospective cohort	Apparent performance	290 consecutive patients with rectal bleeding presenting to 83 GPs in Limburg (the Netherlands) from September 1988 to April 1990. Predictors: questionnaires completed by GPs and patients, and laboratory test results
Hodder 2005[77]	Netherlands model (Fijten 1995[55])	Colorectal	UK	Evaluated in secondary care, model developed in primary care (Fijten 1995[55])	Prospective cohort	External validation	Patients referred from primary care with colorectal symptoms over a 3-year period to the Leighton Hospital, Crewe, UK
Elias 2017[50]	'Netherlands model' (Fijten 1995[55])	Colorectal	The Netherlands	Primary care	Prospective cross-sectional	External validation	CEDAR study: patients referred to endoscopy centres by participating Dutch primary care practices. 2009–12
Kop 2015[61]	No name (machine learning)	Colorectal	The Netherlands	Primary care	Retrospective case–control	Apparent performance	Anonymised electronic records from two general practice database systems in the Utrecht region, the Netherlands, between 1 July 2006 and 31 December 2011
Nørrelund 1996[57]	Danish (Nørrelund) model	Colorectal	Denmark	Primary care	Prospective cohort	Apparent performance II	Patients presenting to GPs with first episode of rectal bleeding • Study 1: 750 GPs, 1989–91 • Study 2: 450 GPs, 1991–2
Elias 2017[50]	Danish (Nørrelund) model	Colorectal	The Netherlands	Primary care	Prospective cross-sectional	External validation	CEDAR study: patients referred to endoscopy centres by participating Dutch primary care practices. 2009–12
Hippisley-Cox 2012[80]	QCancer	Colorectal	UK	Primary care	Open prospective cohort	Internal validation II	QResearch database
Collins 2012[15]	QCancer	Colorectal	UK	Primary care	Retrospective cohort	External validation	THIN database

continued

TABLE 37 Systematic review 2: included studies, study characteristics (1) (*continued*)

Study	Prediction model	Cancer type(s)	Country	Setting	Study design	Stage of development	Data source
Hamilton 2005[81]	RAT	Colorectal	UK	Primary care	Case–control	Apparent performance	Patients attending all 21 general practices in Exeter, UK. Cases identified from the cancer registry at the Royal Devon and Exeter Hospital
Elias 2017[50]	RAT	Colorectal	The Netherlands	Primary care	Prospective cross-sectional	External validation	CEDAR study: patients referred to endoscopy centres by participating Dutch primary care practices. 2009–12
Hamilton 2009[74]	RAT	Colorectal	UK	Primary care	Case–control	Model updating	THIN database
Stapley 2017[72]	RAT	Bowel	UK	Primary care	Case–control	Apparent performance	GPRD (now called the CPRD)
Gastro-oesophageal							
Hippisley-Cox 2011[82]	QCancer	Gastro-oesophageal	UK	Primary care	Open prospective cohort	Internal validation II	QResearch database
Collins 2013[88]	QCancer	Gastro-oesophageal	UK	Primary care	Retrospective cohort	External validation	THIN database
Stapley 2013[71]	RAT	Gastro-oesophageal	UK	Primary care	Case–control	Apparent performance	GPRD (now called the CPRD)
Lung							
Iyen-Omofoman 2013[21]	No name	Lung	UK	Primary care	Case–control	Internal validation II	THIN database
Hippisley-Cox 2011[52]	QCancer	Lung	UK	Primary care	Open prospective cohort	Internal validation II	QResearch database
Hamilton 2005[51]	RAT	Lung	UK	Primary care	Case–control	Apparent performance	Patients attending all 21 general practices in Exeter, UK. Cases identified from the cancer registry at the Royal Devon and Exeter Hospital

Study	Prediction model	Cancer type(s)	Country	Setting	Study design	Stage of development	Data source
Ovarian							
Hippisley-Cox 2011[84]	QCancer	Ovarian	UK	Primary care	Open prospective cohort	Internal validation II	QResearch database
Collins 2013[89]	QCancer	Ovarian	UK	Primary care	Retrospective cohort	External validation	THIN database
Hamilton 2009[83]	RAT	Ovarian	UK	Primary care	Case–control	Apparent performance	Patients attending 39 general practices in Devon, UK
Pancreas							
Hippisley-Cox 2012[32]	QCancer	Pancreas	UK	Primary care	Open prospective cohort	Internal validation II	QResearch database
Collins 2013[14]	QCancer	Pancreas	UK	Primary care	Retrospective cohort	External validation	THIN database
Stapley 2012[70]	RAT	Pancreatic	UK	Primary care	Case–control	Apparent performance	GPRD (now called the CPRD)
Keane 2014[62]	No name	PDAC and BTCs	UK	Primary care	Case–control	Apparent performance	THIN database
Prostate							
Hamilton 2006[90]	RAT	Prostate	UK	Primary care	Case–control	Apparent performance	Patients attending all 21 general practices in Exeter, UK. Cases identified from the cancer registry at the Royal Devon and Exeter Hospital
Renal							
Hippisley-Cox 2012[85]	QCancer	Renal	UK	Primary care	Open prospective cohort	Internal validation II	QResearch database
Collins 2013[91]	QCancer	Renal	UK	Primary care	Retrospective cohort	External validation	THIN database
Shephard 2013[66]	RAT	Renal	UK	Primary care	Case–control	Apparent performance	GPRD (now called the CPRD)

continued

TABLE 37 Systematic review 2: included studies, study characteristics (1) (*continued*)

Study	Prediction model	Cancer type(s)	Country	Setting	Study design	Stage of development	Data source
Uterine							
Walker 2013[73]	RAT	Uterine	UK	Primary care	Case–control	Apparent performance	GPRD (now called the CPRD)
Metastatic							
Hamilton 2015[86]	RAT	Metastatic (breast, colorectal, prostate)	UK	Primary care	Case–control	Apparent performance	Patients attending 11 general practices in Devon, UK
Multiple							
Hippisley-Cox 2013[53]	QCancer	Females – multiple cancers (including breast, cervical, ovarian, uterine, blood, colorectal, gastro-oesophageal, lung, pancreatic, renal)	UK	Primary care	Open prospective cohort	Internal validation II	QResearch database
Hippisley-Cox 2013[54]	QCancer	Males – multiple cancers (including lung, colorectal, gastro-oesophageal, pancreatic, renal tract, blood, prostate, testicular)	UK	Primary care	Open prospective cohort	Internal validation II	QResearch database
Dommett 2013[64]	No name	Multiple paediatric cancers	UK	Primary care	Prospective case–control	Apparent performance	CPRD
Muris 1995[56]	Netherlands (Muris 1995[56]) model	Multiple	The Netherlands	Primary care	Prospective cohort	Apparent performance	Patients presenting to GPs for new abdominal complaints. 1989
Elias 2017[50]	Netherlands (Muris 1995) model	Colorectal, yet Muris 1995 contains multiple cancers	The Netherlands	Primary care	Prospective cross-sectional	External validation	CEDAR study: patients referred to endoscopy centres by participating Dutch primary care practices. 2009–12

BTC, biliary tract cancer; CEDAR, Cost-Effectiveness of a Decision rule for Abdominal complaints in Primary care; GPRD, General Practice Research Database; PDAC, pancreatic ductal adenocarcinoma.

TABLE 38 Systematic review 2: included studies, study characteristics (2)

Study	Participants	Candidate predictors	Outcome to be predicted	Sample size	Number/handling of missing data
Bladder					
Shephard 2012[69]	Inclusion criteria: • Cases – patients with a first diagnosis of bladder cancer between January 2000 and December 2009 inclusive; aged ≥ 40 years; a minimum of 1 year of data before diagnosis. The first instance of a bladder cancer code was assigned the date of diagnosis: the index date. The index date for controls was the index date for their matched case • Controls – up to five controls were matched on sex, general practice, and to 1 year of age of the case Exclusion criteria: metastatic cancer of the bladder from a non-bladder primary, diagnosis before 2000, or no consultations in the year before diagnosis	43 symptoms and 104 abnormal test results were considered initially. However, only six symptoms and seven abnormal test results occurred in at least 5% of cases	Bladder cancer within 1 year of presentation	26,633	NR
Blood					
Shephard 2016[67]	Inclusion criteria: • Cases – aged ≥ 40 years, diagnosed with leukaemia between 2000 and 2009 inclusive • Controls – up to five for each case, matched on age, sex and general practice Exclusion criteria: cases with reticuloendothelial cancer or thrombocytic leukaemia and their matched controls; < 1 year of records before the index date; cases without controls; controls with leukaemia; and controls who had not sought medical care after registration	50 symptoms were considered initially	Leukaemia diagnosis within 1 year of presentation	4655 (2877 chronic, 937 acute) leukaemia cases and 22,852 controls (12,811 for chronic and 4214 for acute)	NR

continued

TABLE 38 Systematic review 2: included studies, study characteristics (2) (*continued*)

Study	Participants	Candidate predictors	Outcome to be predicted	Sample size	Number/handling of missing data
Shephard 2015[68]	Inclusion criteria: • Cases – aged > 40 years with a diagnosis of myeloma between January 2000 and December 2009. Date of diagnosis was the first myeloma code, which served as the index date for the matched controls • Controls – up to five for each case, matched for age, sex and practice Exclusion criteria: < 1 year of records before the index date; cases without controls; controls with myeloma; controls who had not sought medical care after registration	62 symptoms and 22 abnormal test results were considered initially where they occurred in at least 2% of cases or controls	Myeloma within 1 year of presentation	14,860 patients (2703 cases; 12,157 controls)	NR
Brain					
Hamilton 2007[87]	Inclusion criteria: • Cases – patients aged ≥ 18 years with a brain tumour diagnosed between May 1988 and March 2006 • Controls – all potential controls were matched on sex, general practice and age Exclusion criteria: • Cases and controls – < 2 years of data before the first tumour code; no consultations in the 6 months before diagnosis • Controls – previous brain tumours	Libraries of codes for clinical variables previously described with brain tumours were assembled and occurrences of these variables in the 6 months before the index date in cases and controls were identified. Variables were retained only if they occurred in at least 1% of cases or controls	Diagnosis of brain cancer within 1 year of presentation	• Cases: 3505 • Controls: 17,173	NR

Study	Participants	Candidate predictors	Outcome to be predicted	Sample size	Number/handling of missing data
Breast					
McCowan 2011[60]	Derivation cohort: • Inclusion criteria – consecutive patients referred by a GP between February and June 2007 • Exclusion criteria – attending for reasons related to cosmetic surgery or previous conservative breast surgery; patients who were included in the validation cohort Validation cohort: • Inclusion criteria – all women who attended for an initial consultation regarding symptomatic breast problems in 11 participating general practices in the region between January 2006 and June 2007 • Exclusion criteria – consultation related to issues around cosmetic surgery or breastfeeding problems; refusal to participate; loss of contact with the patient	Age, socioeconomic status (as assessed by the Carstairs Index), obesity classification, smoking history, menopausal status, regular periods, current pregnancy, contraception, hormone replacement therapy, hysterectomy, number of pregnancies, past breast problems history, signs after examination, lump characteristics (size, shape, mobility, texture), nipple discharge	Breast cancer diagnosis within 1 year of presentation	• Derivation sample: 802 patients, 59 diagnosed with cancer • Validation sample: 202 patients, five diagnosed with cancer	NR

continued

TABLE 38 Systematic review 2: included studies, study characteristics (2) (*continued*)

Study	Participants	Candidate predictors	Outcome to be predicted	Sample size	Number/handling of missing data
Colorectal					
Marshall 2011[22]	Inclusion criteria: • Cases – patients aged ≥ 30 years with a diagnosis of CRC between January 2001 and July 2006 • Controls – seven controls per case, matched for practice, sex and age Exclusion criteria: cases with < 2 years of data before diagnosis	Constipation episode, laxative prescription, diarrhoea episode, antimotility prescription, change in bowel habit (diarrhoea), change in bowel habit (constipation), change in bowel habit (no diarrhoea/ constipation), IBS diagnosis, antispasmodic prescription, rectal bleeding or melaena, faecal occult blood present, weight loss of > 10% in 2 years, weight loss of 5–10% in 2 years, unknown/no weight loss, abdominal pain/tenderness, abnormal rectal examination, anaemia with low levels of Hb, iron prescription, flatulence, diabetes, BMI of > 30 kg/m^2, DVT/PE, abdominal mass, mean cell volume of < 80 fl	Diagnosis of CRC	Derivation: cases, 5477; controls, 38314 Discrimination based on two data sets: derivation (THIN) data set and CAPER data set (349 cases; 1744 controls)	NR
Elias 2017[50]	Inclusion criteria: lower abdominal complaints for at least 2 weeks with rectal bleeding, change in bowel habit, abdominal pain, fever, diarrhoea, weight loss, sudden onset in elderly, and/or physical examination suggestive of colorectal disease	N/A. Used BB equation (Marshall 2011[22])	CRC	810, including 37 with CRC	Multiple imputation

Study	Participants	Candidate predictors	Outcome to be predicted	Sample size	Number/handling of missing data
Fijten 1995[55]	Inclusion criteria: overt rectal bleeding reason for GP visit, or history of recent rectal blood loss (within the previous 3 months) Exclusion criteria: patients aged < 18 or > 75 years, pregnant, urgent admission to a hospital (e.g. for a massive bleeding or acute abdominal pain), follow-up data not available	Relevant items were identified from literature. Data collected by: 1. GP questionnaire after the first consultation, containing 70 variables; history, physical examination and initial management 2. Patient questionnaire, containing 150 (somatic and psychological) questions 3. Laboratory tests: Hb levels, ESR (several cut-off points), WBC and occult blood in the faeces	Diagnosis of CRC within at least 1 year of study entry, from medical records and GP	269 patients, nine cancer diagnoses	• Missing laboratory test data led to excluding these data in the logistic regression • 21 patients lost to follow-up • Age and sex of these patients no different from those included
Hodder 2005[77]	All patients referred from primary care with colorectal symptoms from October 1999 to October 2002	N/A. Used Fijten 1995[55]	Diagnosis of CRC	3302 patients, 156 diagnosed with cancer	NR
Elias 2017[50]	Inclusion criteria: lower-abdominal complaints for at least 2 weeks with rectal bleeding, change in bowel habit, abdominal pain, fever, diarrhoea, weight loss, sudden onset in elderly, and/or physical examination suggestive of colorectal disease	N/A. Used Fijten 1995[55]	CRC	810, including 37 with CRC	Multiple imputation

continued

TABLE 38 Systematic review 2: included studies, study characteristics (2) (*continued*)

Study	Participants	Candidate predictors	Outcome to be predicted	Sample size	Number/handling of missing data
Kop 2015[61]	Patients aged ≥ 30 years • Cases: a 6-month period preceding the diagnosis • Controls: a 6-month period was randomly chosen between 1 July 2006 and 31 December 2011	Age; sex; number of times each of 720 ICPC codes were recorded; number of times each of 86 medication (ATC) code was prescribed; number of times patient was referred to a specialist for each of the 94 referral codes; 33 attributes, each representing whether or not a patient is currently in a certain (medical) condition; number of times Hb was inside/outside threshold; number of times a patient's cell volume value was inside/outside the desired range; number of times blood was found/was not found in a patient's stool; whether or not patient's medical history contains a certain temporal pattern	Diagnostic of CRC	808 CRC diagnoses occurring in 219,447 patients	NR
Nørrelund 1996[57]	Inclusion criteria: aged ≥ 40 years, presenting with first episode of rectal bleeding within previous 6 months Study 1: 1989–91; study 2: 1991–2 Exclusion criteria: known inflammatory bowel disease, colonic polyps, polyposis coli, CRC, predisposition to haemorrhage (e.g. coagulation defect), melaena stool Conducted follow-up in 1994: 0 (study 1) and 2 (study 2) additional diagnoses of CRC observed	• Weight loss, abdominal pain, changes in bowel habits, discomfort • Patient perceptions of cause of bleeding • Age, sex	CRC	• Study 1: 208, including 32 with CRC • Study 2: 209, including 22 with CRC	NR

Study	Participants	Candidate predictors	Outcome to be predicted	Sample size	Number/handling of missing data
Elias 2017[50]	Inclusion criteria: lower abdominal complaints for at least 2 weeks with rectal bleeding, change in bowel habit, abdominal pain, fever, diarrhoea, weight loss, sudden onset in elderly, and/or physical examination suggestive of colorectal disease	N/A. Used Danish (i.e. Nørrelund 1996[57]) model	CRC	810, including 37 with CRC	Multiple imputation
Hippisley-Cox 2012[80]	Inclusion criteria: patients aged 30–84 years registered between 1 January 2000 and 30 September 2010 Exclusion criteria: missing a postcode-related Townsend deprivation score; history of pancreatic cancer at baseline; red-flag symptoms recorded in the 12 months before the study entry date Inclusion criteria: patients from general practices using EMIS computer system for a minimum of 1 year	• Current first onset of rectal bleeding, loss of appetite, weight-loss symptom or abdominal pain • First onset (within the previous 12 months) of abdominal distension, constipation, diarrhoea, change in bowel habit or tiredness • Age, BMI, alcohol status, smoking status, Townsend deprivation score (postcodes), family history of gastrointestinal cancer, previous diagnosis of cancer apart from CRC, IBD (Crohn's disease, ulcerative colitis, coeliac disease), previous history of gastrointestinal polyp, diabetes, anaemia	Diagnosis of CRC, defined as incident diagnosis of CRC during the 2 years after study	• Derivation sample: 2,351,052 • Validation sample: 1,236,601	Multiple imputation to replace missing values for BMI, alcohol intake and smoking status
Collins 2012[15]	• Inclusion criteria: patients registered between 1 January 2000 and 30 June 2008, and recorded on THIN database • Exclusion criteria: prior diagnosis of CRC, patients registered < 12 months with the general practice, had invalid dates, aged < 30 years or > 85 years	N/A. Used QCancer (colorectal); see Hippisley-Cox 2012[80]	Diagnosis of CRC, defined as incident diagnosis of CRC during the 2 years after study entry	2,135,540 patients, including 3712 cases of CRC (1676 women and 2036 men)	Multiple imputation using all predictors plus the outcome variable was used to replace missing values for alcohol consumption

continued

TABLE 38 Systematic review 2: included studies, study characteristics (2) (*continued*)

Study	Participants	Candidate predictors	Outcome to be predicted	Sample size	Number/handling of missing data
Hamilton 2005[81]	Inclusion criteria: • Cases – patients aged ≥ 40 years with a primary CRC, diagnosed between 1998 and 2002 in one hospital • Controls: five for each case matched on sex, general practice and age Exclusion criteria (cases and controls): unobtainable records; no consultations in the 2 years before diagnosis; previous CRC; or residence outside Exeter at the time of diagnosis	All related features (symptoms and variables) were investigated; only features occurring in at least 2.5% of either cases or controls were analysed	Diagnosis of CRC within 2 years of presentation	349 cases; 1744 controls	NR
Elias 2017[50]	Inclusion criteria: lower-abdominal complaints for at least 2 weeks with rectal bleeding, change in bowel habit, abdominal pain, fever, diarrhoea, weight loss, sudden onset in elderly, and/or physical examination suggestive of colorectal disease	N/A. Used RAT (colorectal)	CRC	810, including 37 with CRC	Multiple imputation
Hamilton 2009[74]	Inclusion criteria: • Cases – all patients with CRC, aged ≥ 30 years, diagnosed between January 2001 and July 2006 • Controls – up to seven for each case, matched for practice, sex and age Exclusion criteria: < 2 years of data recorded prior to the diagnosis	23 candidate variables (features) were identified from a review of the literature	Diagnosis of CRC within 2 years of presentation	5477 cases; 38,314 controls	NR

Study	Participants	Candidate predictors	Outcome to be predicted	Sample size	Number/handling of missing data
Stapley 2017[72]	Inclusion criteria: • Cases – patients with IBD and CRC, aged 18–49 years at diagnosis, between January 2000 and December 2013 • Controls – three controls were matched on sex, general practice, and to 1 year of age of the case Exclusion criteria: cases and controls with no consultations in the year before the index date; controls that had a previous diagnostic code of CRC/IBD, diagnosis before 2000	All symptoms, physical signs or abnormal investigations related to CRC/IBD that occurred in ≥ 5% of cases or controls were retained	Diagnosis of either IBD or CRC within 1 year of presentation with symptoms	• CRC: 1661 cases; 3979 controls • IBD: 9578 cases; 22,947 controls	NR
Gastro-oesophageal					
Hippisley-Cox 2011[82]	Inclusion criteria: patients aged 30–84 years, registered with practices between 1 January 2000 and 30 September 2010 Exclusion criteria: patients without a postcode-related Townsend deprivation score, history of gastro-oesophageal cancer at baseline, red-flag symptoms recorded (dysphagia, haematemesis, loss of appetite, weight loss and abdominal pain) in the 12 months prior to the study entry date Inclusion criteria: patients from general practices using EMIS computer system for a minimum of 1 year	Current first onset of dysphagia, haematemesis, loss of appetite, weight-loss symptom or abdominal pain; recent GP consultation for tiredness; age; BMI; smoking status; alcohol status; Townsend deprivation score; family history of gastrointestinal cancer; previous diagnosis of cancer apart from gastro-oesophageal cancer; anaemia	Diagnosis of either gastric cancer or oesophageal cancer in the 2 years following presentation	• Derivation sample: 2,355,719 • Validation sample: 1,238,971	Multiple imputation to replace missing values for BMI, alcohol intake and smoking status
Collins 2013[88]	Inclusion criteria: patients registered between 1 January 2000 and 30 June 2008, and recorded on THIN database Exclusion criteria: prior diagnosis of CRC, patients registered < 12 months with the general practice, had invalid dates, aged < 30 or > 85 years	N/A. Used QCancer (gastro-oesophageal), see Hippisley-Cox 2011[52]	Diagnosis of gastro-oesophageal cancer in the 2 years after study entry	2,135,540 patients, including 1766 cases of gastro-oesophageal cancer	Multiple imputation using all predictors plus the outcome variable was used to replace missing values for alcohol consumption

continued

TABLE 38 Systematic review 2: included studies, study characteristics (2) (*continued*)

Study	Participants	Candidate predictors	Outcome to be predicted	Sample size	Number/handling of missing data
Stapley 2013[71]	Inclusion criteria: • Cases: patients aged ≥ 40 years with an oesophago-gastric tumour diagnosed between 1 January 2000 and 31 December 2009, and with at least 1 year of data • Controls: up to five for each case, matched by year of birth, sex and practice Exclusion criteria: any individuals without a consultation in the year before the diagnosis/index date; cancers from other sites that had spread to the oesophagus or stomach; controls if they had ever had oesophagogastric cancer	All symptoms, physical signs or abnormal investigations that occurred in ≥ 5% of cases or controls were retained	Gastro-oesophageal cancer diagnosis within 1 year of presentation	Oesophageal cancer: 4854 cases, 21,506 controls Gastric cancer: 2617 cases, 11,371 controls	NR
Lung					
Iyen-Omofoman 2013[21]	Inclusion criteria: • Cases – all cancers diagnosed between 1 January 2000 and 28 July 2009; aged ≥ 40 years; a minimum of 1 year of data before diagnosis • Controls – matched on age, sex and general practice	• All symptoms, diagnoses and investigations in 2 years prior to lung cancer diagnosis/index date • Age, sex, socioeconomic status, smoking history, haemoptysis, cough, chest/shoulder pain, dyspnoea, weight loss, hoarseness, finger clubbing, features suggestive of metastasis from lung cancer, cervical/supraclavicular lymphadenopathy, upper and lower respiratory tract infections, non-specific chest infections, constipation, depressive disorders, chronic obstructive pulmonary disease • Assessed records of chest X-rays, blood tests and other GP consultations	Diagnosis of lung cancer within 12 months of presentation	12,074 cases; 120,731 controls	Unclear

Study	Participants	Candidate predictors	Outcome to be predicted	Sample size	Number/handling of missing data
Hippisley-Cox 2011[52]	• Inclusion criteria: patients aged 30–84 years, drawn from patients registered with practices between 1 January 2000 and 30 September 2010 • Exclusion criteria: missing a postcode-related Townsend deprivation score; history of lung cancer at baseline; red-flag symptoms (haemoptysis, loss of appetite, or weight loss) recorded in the 12 months before the study entry date • Inclusion criteria: patients from general practices using EMIS computer system for a minimum of 1 year	Current first onset of haemoptysis, loss of appetite or weight-loss symptom; recent (within 12 months) GP consultation for cough, dyspnoea, tiredness or hoarseness; BMI; smoking status; chronic obstructive airways disease diagnosed ever; Townsend deprivation score; family history of lung cancer; previous diagnosis of cancer apart from lung cancer; asthma diagnosed ever; pneumonia diagnosed ever; asbestos exposure ever; anaemia	Diagnosis of lung cancer during the 2 years after study entry	• Derivation sample 2,406,127 • Validation sample 1,342,329	Multiple imputation to replace missing values for smoking and BMI
Hamilton 2005[51]	Inclusion criteria: • Cases – residents of Exeter, UK, aged ≥ 40 years who had a primary lung cancer diagnosed during 1998–2002, inclusive • Controls – five for each case matched on sex, general practice, and age; alive at time of case diagnosis Exclusion criteria (both): unobtainable records; no consultations in the 2 years before diagnosis; previous lung cancer; or residence outside Exeter at the time of diagnosis	All related features (symptoms and variables) were investigated. Only features occurring in at least 2.5% of either cases or controls were analysed	Diagnosis of lung cancer within 2 years of presentation	247 cases, 1235 controls	NR

continued

TABLE 38 Systematic review 2: included studies, study characteristics (2) (*continued*)

Study	Participants	Candidate predictors	Outcome to be predicted	Sample size	Number/handling of missing data
Ovarian					
Hippisley-Cox 2011[84]	• Inclusion criteria: patients aged 30–84 years registered between 1 January 2000 and 30 September 2010 • Exclusion criteria: missing a postcode-related Townsend deprivation score; history of bilateral oophorectomy or ovarian cancer at baseline; red-flag symptoms recorded in the 12 months before the study entry date, aged < 30 or > 84 years	• Current first onset of loss of appetite, weight loss symptom, abdominal pain, abdominal distension, rectal bleeding or postmenopausal bleeding • Recent (within 12 months) GP consultation for constipation, diarrhoea, tiredness and urinary frequency • Age, BMI, smoking status, alcohol use, Townsend deprivation score, previous diagnosis of cancer apart from ovarian cancer, anaemia	Diagnosis of ovarian cancer during the 2 years after study entry	• Derivation sample: 1,158,723 • Validation sample: 608,862	Multiple imputation to replace missing values for BMI, alcohol intake and smoking status
Collins 2013[89]	• Inclusion criteria: patients registered between 1 January 2000 and 30 June 2008 on THIN database • Exclusion criteria: missing a postcode-related Townsend deprivation score; history of bilateral oophorectomy or ovarian cancer at baseline; red-flag symptoms recorded in the 12 months before the study entry date; aged < 30 or > 84 years	N/A. Used QCancer (ovarian), see Hippisley-Cox 2012[84]	Diagnosis of ovarian cancer during the 2 years after study entry	1,054,818 patients (1,827,997 person-years)	No missing data reported
Hamilton 2009[83]	Inclusion criteria: • Cases – all women with primary ovarian cancer, aged ≥ 30 years, diagnosed between 2000 and 2007 • Controls – five for each case, matched for practice, sex, and age; alive at the time of diagnosis of their matched case Exclusion criteria: medical record unobtainable, no entry in the records in the year before diagnosis, previous ovarian cancer or bilateral oophorectomy, or residence outside the study area at the time of diagnosis	Only symptoms occurring in at least 5% of either cases or controls were studied	Diagnosis of ovarian cancer within 1 year of presentation	212 cases; 1060 controls	• NR • 11 practices declined to participate

Study	Participants	Candidate predictors	Outcome to be predicted	Sample size	Number/handling of missing data
Pancreas					
Hippisley-Cox 2012[32]	Inclusion criteria: patients aged 30–84 years registered between 1 January 2000 and 30 September 2010; missing a postcode-related Townsend deprivation score; history of pancreatic cancer at baseline; red-flag symptoms recorded in the 12 months before the study entry date, aged < 30 or > 84 years	• Current first onset of dysphagia, loss of appetite, weight loss symptom, abdominal pain or abdominal distension • Recent (within 12 months) GP consultation for constipation, diarrhoea, tiredness, itching • Age, BMI, smoking status, alcohol status, Townsend deprivation score, derived from patients' postcodes, diabetes, pancreatitis (acute/chronic/none) at study entry, previous diagnosis of cancer apart from pancreatic cancer at study entry, anaemia	Presence of pancreatic cancer	• Derivation sample: 2,364,571 • Validation sample: 1,243,740	Multiple imputation to replace missing values for BMI, alcohol intake and smoking status
Collins 2013[14]	• Inclusion criteria: patients registered between 1 January 2000 and 30 June 2008 on THIN database • Exclusion criteria: missing data; history of pancreatic cancer at baseline; red-flag symptoms recorded in the 12 months before the study entry date, aged < 30 or > 84 years	N/A. Used QCancer (pancreas), see Hippisley-Cox 2012[32]	Diagnosis of pancreatic cancer within 2 years of study entry	2,150,322 patients (3,744,567 person-years)	Multiple imputation using all predictors plus the outcome variable was used to replace missing values for smoking status
Stapley 2012[70]	Inclusion criteria: • Cases – patients aged ≥ 40 years with a pancreatic tumour diagnosed between 1 January 2000 and 31 December 2009, with at least 1 year of data before the first diagnostic code • Controls – up to five controls, matched to the case by year of birth, sex and practice Exclusion criteria: those with no consultations in the year before cancer diagnosis/index date; controls excluded if ever had pancreatic cancer	All symptoms, physical signs or abnormal investigations that occurred in ≥ 5% of cases or controls were retained	Pancreatic cancer diagnosis within 1 year of presentation	3635 cases, 16,459 controls	NR

continued

TABLE 38 Systematic review 2: included studies, study characteristics (2) (*continued*)

Study	Participants	Candidate predictors	Outcome to be predicted	Sample size	Number/handling of missing data
Keane 2014[62]	• Cases: patients with a Read code diagnosis of PDAC or BTC between 1 January 2000 and 31 December 2010 • Controls: no PDAC or BTC diagnosis, matched to cases on age, sex, practice and year of diagnosis (random consultation date) • Two years of data available prior to the index date • Only included general practices that had acceptable mortality recording and computer usage	'Alarm' symptoms and commonly performed blood test results were identified based on clinical knowledge and the existing literature. Only symptoms with a frequency of > 5% were identified as potential alarm symptoms. Age, sex, time period and Townsend deprivation score, smoking status and BMI were selected as potential confounders	Diagnosis of PDAC or BTC within 2 years of presentation	2773 cases with PDAC, 848 cases with BTC and 15,395 controls	NR
Prostate					
Hamilton 2006[90]	Inclusion criteria: • Cases – residents of Exeter, UK aged ≥ 40 years who had a prostate cancer diagnosed from 1998 to 2002, inclusive • Controls – five for each case matched on sex, general practice, and age; alive at time of index date (case diagnosis date) Exclusion criteria (both): unobtainable records; no consultations in the 2 years before diagnosis; previous prostate cancer; or residence outside Exeter at the time of diagnosis	All related features (symptoms and variables) were investigated; only features occurring in at least 2.5% of either cases or controls were analysed	Diagnosis of prostate cancer within 2 years of presentation	217 cases, 1080 controls	NR

Study	Participants	Candidate predictors	Outcome to be predicted	Sample size	Number/handling of missing data
Renal					
Hippisley-Cox 2012[85]	• Inclusion criteria: patients aged 30–84 years registered between 1 January 2000 and 30 September 2010 • Exclusion criteria: missing a postcode-related Townsend deprivation score; history of renal tract cancer at baseline; red-flag symptoms recorded in the 12 months before the study entry date, aged < 30 or > 84 years	• Current first onset of macroscopic haematuria, loss of appetite, weight-loss symptom or abdominal pain • Recent (within 12 months) GP consultation for constipation, diarrhoea or tiredness • Age, BMI, smoking status, alcohol use, Townsend deprivation score, derived from patients' postcodes (continuous), treated hypertension, renal stones, structural kidney problems, diabetes (type 1/type 2/no diabetes), previous diagnosis of cancer apart from renal tract cancer at study entry, anaemia	Diagnostic of renal tract cancer within 2 years of presentation	• Derivation sample: 2,359,168 • Validation sample: 1,240,722	Multiple imputation to replace missing values for BMI, alcohol intake and smoking status
Collins 2013[91]	• Inclusion criteria: patients registered between 1 January 2000 and 30 June 2008 on THIN database • Exclusion criteria: missing data; history of renal cancer at baseline; red-flag symptoms recorded in the 12 months before the study entry date, aged < 30 or > 84 years	N/A. Used QCancer (renal), see Hippisley-Cox 2012[32]	Presence of renal cancer	2,145,133 patients (3,731,312 person-years)	Multiple imputation using all predictors plus the outcome variable was used to replace missing values for smoking status

continued

TABLE 38 Systematic review 2: included studies, study characteristics (2) (*continued*)

Study	Participants	Candidate predictors	Outcome to be predicted	Sample size	Number/handling of missing data
Shephard 2013[66]	Inclusion criteria: • Cases – patients with a first diagnosis of kidney cancer between January 2000 and December 2009, inclusive; aged ≥ 40 years; a minimum of 1 year of data before diagnosis • Controls: up to five controls were matched on sex, general practice, and to 1 year of age of the case Exclusion criteria: metastatic cancer of the kidney from a non-kidney primary, diagnosis before 2000, or no consultations in the year before diagnosis	• Features potentially associated with kidney cancer were reviewed from a traditional literature review, patient-reported symptoms and clinical knowledge • Only those reported in > 5% of cases or controls were considered further	Diagnosis of kidney cancer within 1 year of presentation	17,240 patients (3149 cases; 14,091 controls)	NR
Uterine					
Walker 2013[73]	Inclusion criteria: • Cases – women aged ≥ 40 years with a diagnostic code for uterine tumour, between 1 January 2000 and 31 December 2009 • Controls: five controls were matched on sex, general practice, and to 1 year of age of the case Exclusion criteria: leiomyosarcoma; diagnosis before 1 January 2000; controls diagnosed with uterine cancer before the index date; metastatic cancer from a non-uterine primary cancer; women with a recorded hysterectomy before the index date; and women with no consultations in the year before the index date	All symptoms, physical signs or abnormal investigations related to uterine cancer, that occurred in ≥ 2% of cases or controls were retained	Diagnosis of uterine tumour within 1 year of presentation with symptoms	2732 cases, 9537 controls	NR

Study	Participants	Candidate predictors	Outcome to be predicted	Sample size	Number/handling of missing data
Metastatic					
Hamilton 2015[86]	Inclusion criteria: • Cases – deceased with a prior record of breast, colorectal or prostate cancer and radiologically or histologically proven metastatic cancer • Controls – two controls per case, alive at the time of the diagnosis of metastatic cancer in the case, matched for practice, sex, age; one with same (non-metastatic) cancer, one without cancer Exclusion criteria: primary cancer considered incurable at the time of initial diagnosis, or metastatic spread had occurred within 6 months of diagnosis of the primary cancer; cases for whom the primary cancer was diagnosed before 40 years of age; patients whose metastases occurred before registration at the current practice; cases, the full record for which had been archived off site after death	NR. 207 separate 'features' were identified in > 2% of cases	Diagnosis of metastatic cancer within 6 months of diagnosis of primary cancer	• Cases: 162 • Cancer controls: 152 • Non-cancer controls: 145	NR

continued

TABLE 38 Systematic review 2: included studies, study characteristics (2) (*continued*)

Study	Participants	Candidate predictors	Outcome to be predicted	Sample size	Number/handling of missing data
Multiple					
Hippisley-Cox 2013[53]	• Inclusion criteria: females aged 25–89 years registered with participating practices between 1 January 2000 and 1 April 2012 • Exclusion criteria: missing a postcode-related Townsend deprivation score; red-flag symptoms recorded in the 12 months before the study entry date • Inclusion criteria: patients from general practices using EMIS computer system for a minimum of 1 year	Red-flag symptoms and more general symptoms, plus risk factors: age, BMI, smoking status, alcohol use, Townsend deprivation score (postcodes), previous diagnosis of cancer, anaemia, family history of breast cancer, family history of gastrointestinal cancer, family history of ovarian cancer, benign breast disease, chronic pancreatitis, type 1 diabetes, type 2 diabetes, endometriosis, endometrial hyperplasia or polyp, fibroid, polycystic ovarian disease, rheumatoid arthritis, systemic lupus erythematosus, a HIV infection or AIDS, oral contraceptive use, hormone replacement therapy	Diagnosis of cancer within 2 years after study entry	• Derivation sample: 1,240,864 • Validation sample: 667,603	Multiple imputation was used in the validation cohort to replace missing values for BMI, alcohol, and smoking
Hippisley-Cox 2013[54]	• Inclusion criteria: males aged 25–89 years registered with participating practices between 1 January 2000 and 1 April 2012 • Exclusion criteria: missing a postcode-related Townsend deprivation score; red-flag symptoms recorded in the 12 months before the study entry date • Inclusion criteria: patients from general practices using EMIS computer system for a minimum of 1 year	Red-flag symptoms and more general symptoms, plus risk factors: age, BMI, smoking status, alcohol use, Townsend deprivation score (postcodes), previous diagnosis of cancer, anaemia, family history of gastrointestinal cancer, family history of prostate cancer, chronic pancreatitis, type 1 diabetes, type 2 diabetes	Diagnosis of cancer within 2 years after study entry	• Derivation sample: 1,263,071 • Validation sample: 679,174	Multiple imputation was used in the validation cohort to replace missing values for BMI, alcohol intake and smoking status

Study	Participants	Candidate predictors	Outcome to be predicted	Sample size	Number/handling of missing data
Dommett 2013[64]	Inclusion criteria: • Cases – all children aged 0–14 years, with cancers diagnosed between 1 January 1988 and 31 December 2010 from > 600 general practices across the UK • Controls – up to 13 controls per case registered with the practice on the index date of the case, never diagnosed with cancer, matched on age (within 1 year), sex and practice	Three or more consultations, upper respiratory tract infection, musculoskeletal symptoms, vomiting, cough, headache, lymphadenopathy, rash, abdominal pain, childhood infection, fever, abnormal movement, abdominal mass, pain, fatigue, lump mass swelling (below neck excluding abdomen), eye swelling, shortness of breath, bruising, pallor, bleeding, lump mass swelling of head and neck, visual symptoms, constipation	GPRD medical code for cancer (leukaemia/ lymphoma, central nervous system tumours, bone tumour/soft tissue sarcoma, abdominal tumour) within 1 year of recorded symptom	1267 cases; 15,318 controls	NR
Muris 1995[56]	Inclusion criteria: aged 18–75 years, consulting GP for new abdominal complaints lasting at least 2 weeks, consenting to participate	Age and sex. WBC count, ESR, low Hb level, positive FOBT. Low somatisation score, no depression, high self-esteem, social inadequacy, plus 23 symptoms	Multiple cancers	933, including ≈ 18 having cancer	NR
Elias 2017[50]	Inclusion criteria: lower-abdominal complaints for at least 2 weeks with rectal bleeding, change in bowel habit, abdominal pain, fever, diarrhoea, weight loss, sudden onset in elderly, and/or physical examination suggestive of colorectal disease	N/A. Used Muris 1995[56]	CRC, yet Muris 1995[56] is multiple cancers	810, including 37 with CRC	NR

AIDS, acquired immunodeficiency syndrome; ATC, Anatomical Therapeutic Chemical; BMI, body mass index; BTC, biliary tract cancer; DVT, deep-vein thrombosis; ESR, erythrocyte sedimentation rate; GPRD, General Practice Research Database; HIV, human immunodeficiency virus; IBD, inflammatory bowel disease; IBS, irritable bowel syndrome; ICPC, International Classification of Primary Care; N/A, not applicable; NR, not reported; PDAC, pancreatic ductal adenocarcinoma; PE, pulmonary embolism; WBC, white blood cell.

TABLE 39 Systematic review 2: included studies, model development and performance

Study	Model development	Model performance
Bladder		
Shephard 2012[69]	Conditional logistic regression. Variables with an independent association with cancer ($p < 0.001$) were included in the final multivariable model. Variables were then grouped for multivariable analysis, collecting together variables that were similar (such as visible and non-visible haematuria), using a p-value threshold of ≤ 0.05. The final stage of multivariable analysis used all variables surviving the previous stages, and used a p-value threshold of 0.01. All excluded variables were checked against the final model	PPVs were estimated for predictors shown to be independently associated with cancer in the multivariable analysis. This was repeated for pairs of symptoms and for second attendances with the same symptom (see figure 2 of publication[69])
Blood		
Shephard 2016[67]	Conditional logistic regression. Variables with $p < 0.1$ from univariable analyses were grouped by similarity, those with $p < 0.05$ entered final model, at which point those with $p < 0.01$ were retained in model	• PPVs were calculated for those aged > 60 years, targeting patients around the average age of leukaemia diagnosis • Chronic leukaemia highest PPV was 0.34% for lymphadenopathy. Lymphadenopathy with cough PPV of 0.27%. Many combinations were too rare for PPV calculation • Acute leukaemia highest combined PPV was fever with infection, PPV 0.13%
Shephard 2015[68]	Non-parametric methods in univariable analysis, a p-value threshold of ≤ 0.1 was used to identify candidate variables for multivariable analysis These were then grouped into small clinically coherent groups containing similar variables in the first stage of multivariable analysis, with retention requiring a p-value of ≤ 0.05. A final multivariable model used the surviving variables from the previous stages, using a p-value threshold of 0.01	PPVs were calculated (see figure 2 of publication[68])
Brain		
Hamilton 2007[87]	Variables associated with tumours in univariable analyses with $p \leq 0.1$ were entered into the staged multivariable analyses; discarded variables were checked against the final model. Seven clinically plausible internal interactions were tested in the final model	PPVs: • Headache: 0.09% (95% CI 0.08% to 0.10%) • Motor loss: 0.026% (95% CI 0.024% to 0.030%) • New-onset seizure: 1.2% (95% CI 1.0% to 1.4%) • Confusion: 0.20% (95% CI 0.16% to 0.24%) • Weakness: 0.14% (95% CI 0.11% to 0.18%) • Memory loss: 0.036% (95% CI 0.026% to 0.052%) • Visual disorder: 0.036% (95% CI 0.025% to 0.051%)
Breast		
McCowan 2011[60]	Logistic regression. Univariable associations for the explanatory variables were investigated and those with a threshold of < 0.01 were used in the multivariate regression model	The regression coefficients from the derivation cohort were applied to individuals in the validation cohort and used to generate expected and observed probabilities of breast cancer. The numbers of expected vs. observed cancers were also plotted by decile of predicted risk, based on the derivation model. Hosmer–Lemeshow goodness-of-fit test for the calibration of the model showed no significant difference between expected and observed breast cancers (7.02, $p = 0.73$), but the plot suggested that the number of cancers was overestimated for those at highest risk

TABLE 39 Systematic review 2: included studies, model development and performance (*continued*)

Study	Model development	Model performance
Colorectal		
Marshall 2011[22]	Multivariable conditional logistic regression analysis. Initial univariable conditional logistic regression analysis was carried out with the initial predictor variables and some variables were combined. Variables associated with CRC with $p < 0.1$ were entered into multivariable conditional logistic regression	ROC curves, sensitivity, LRs and PPVs also reported In the CAPER data set, AUCs: BB 0.92 (95% CI 0.91 to 0.94); CAPER 0.91 (95% CI 0.89 to 0.93); NICE guidelines 0.75 (95% CI 0.72 to 0.79)
Elias 2017[50]	N/A. Used BB equation[22]	• NPV: 100 (95% CI 98 to 100) • PPV: 7 (95% CI 5 to 9) • Sensitivity: 97 (95% CI 85 to 100) • Specificity: 36 (95% CI 32 to 40) • AUC: 0.84 (95% CI 0.77 to 0.90)
Fijten 1995[55]	Forward stepwise logistic regression analysis. Variables showing an association ($p < 0.1$) with cancer were included in the multivariate model	• Diagnostic index: AUC 0.97 (no estimate of variance reported). Also reported sensitivity, specificity, predictive values, ORs • Diagnostic index + presence of polyps: AUC 0.92 • For both, calculated cut-off points to maximise sensitivity and specificity
Hodder 2005[77]	Used Fijten 1995[55] risk prediction model. Also assessed performance of guidelines and scores that were not based on risk prediction models (results not reported here)	AUC of 0.775 (SE 0.02)
Elias 2017[50]	N/A. Used Fijten 1995[55]	• NPV: 99 (95% CI 95 to 1000) • PPV: 10 (95% CI 7 to 14) • Sensitivity: 98 (95% CI 83 to 100) • Specificity: 29 (95% CI 24 to 34) • AUC: 0.72 (95% CI 0.62 to 0.81)
Kop 2015[61]	Models were generates using logistic regression, RF, SVM and the CART algorithm	A priori algorithm: the records are scanned first to create frequent patterns of size 1. These patterns are used to generate successively larger patterns, that is k patterns are used to obtain frequent k + 1 patterns. Generating a k + 1 pattern is, however, more elaborate than for standard a priori, because the generated candidate patterns need to accommodate for multiple possible (successions) and (occuring) relations. This results in more k + 1 patterns being tested for frequency than with standard a priori, increasing the complexity
Hippisley-Cox 2012[80]	• Cox proportional hazards. Rubin's rules were used to combine the results across the imputed data sets • Fractional polynomials used for non-linear risk relationships. Fitted full model, variables retained if HR < 0.80, HR > 1.20 and a $p < 0.01$	Females: • R^2 (%): 64.8 (95% CI 63.2 to 66.3) • *D*-statistic: 2.78 (95% CI 2.68 to 2.87) • AUC: 0.89 (95% CI 0.88 to 0.90) Males: • R^2 (%): 66.7 (95% CI 65.3 to 68.0) • *D*-statistic: 2.90 (95% CI 2.81 to 2.98) • AUC: 0.906 (95% CI 0.899 to 0.913)
Collins 2012[15]	N/A. Used QCancer (colorectal), see Hippisley-Cox 2012[80]	Men: Multiple imputation ($m = 10$) ($n = 1{,}059{,}765$) • R^2 (%): 68.32 (95% CI 67.32 to 69.32) • *D*-statistic: 3.00 (95% CI 2.93 to 3.07) • *c*-statistic: 0.918 (95% CI 0.913 to 0.923)

continued

TABLE 39 Systematic review 2: included studies, model development and performance (*continued*)

Study	Model development	Model performance
		Complete case ($n = 417,560$) • R^2 (%): 65.30 (95% CI 63.71 to 66.89) • *D*-statistic: 2.81 (95% CI 2.71 to 2.91) • *c*-statistic: 0.901 (95% CI 0.892 to 0.910) Women: Complete case ($n = 1,075,775$) • R^2 (%): 65.81 (95% CI 64.62 to 67.01) • *D*-statistic (95% CI) 2.84 (95% CI 2.76 to 2.92) • *c*-statistic (95% CI) 0.909 (95% CI 0.903 to 0.915)
Nørrelund 1996[57]	Mann–Whitney and chi-squared tests, and logistic regression. No further details reported	Study 1 (apparent performance I): • Age > 69 years: sensitivity 75%, specificity 76%, PPV 36%, NPV 94% • Age > 69 years plus change in bowel habits: sensitivity 44%, specificity 94%, PPV 56%, NPV 90% • Change in bowel habits plus patient belief bleeding is due to cancer: sensitivity 22%, specificity 97%, PPV 58%, NPV 87% Study 2 (apparent performance II), new bleeders: • Age > 69 years: sensitivity 46%, specificity 72%, PPV 18%, NPV 91% • Age > 69 years plus change in bowel habits: sensitivity 15%, specificity 88%, PPV 13%, NPV 85% • Change in bowel habits plus patient belief bleeding is due to cancer: sensitivity 0%, specificity 95%, PPV 0%, NPV 87% Study 2 (apparent performance II), new or changed bleeders: • Aged > 69 years: sensitivity 45%, specificity 75%, PPV 23%, NPV 91% • Aged > 69 years plus change in bowel habits: sensitivity 23%, specificity 88%, PPV 24%, NPV 87% • Change in bowel habits plus patient belief bleeding is due to cancer: sensitivity 5%, specificity 96%, PPV 14%, NPV 86%
Elias 2017[50]	N/A. Used Danish (Nørrelund 1996[57]) model	• NPV: 93 (95% CI 71 to 99) • PPV: 8 (95% CI 6 to 12) • Sensitivity: 95 (95% CI 80 to 99) • Specificity: 6 (95% CI 3 to 9) • AUC: 0.60 (95% CI 0.48 to 0.72)
Hamilton 2005[81]	Variables associated with cancer in univariable analyses, using a *p*-value of ≤ 0.1, entered the multivariable analysis. In the first stage, similar variables were grouped together; in the second stage, analyses were repeated with the new groups	Nine features [constipation, diarrhoea, rectal bleeding, loss of weight, abdominal pain, abdominal tenderness, abnormal rectal examination, anaemia (Hb level 10–13 g/dl; Hb < 10 g/dl) and a blood sugar concentration of > 10 mmol/l] were associated with CRC before diagnosis. The PPVs (95% CI%) of these were rectal bleeding 2.4% (1.9% to 3.2%), weight loss 1.2% (0.91% to 1.6%), abdominal pain 1.1% (0.86% to 1.3%), diarrhoea 0.94% (0.73% to 1.1%), constipation 0.42% (0.34% to 0.52%), abnormal rectal examination 4.0% (2.4% to 7.4%), abdominal tenderness 1.1% (0.77% to 1.5%), a Hb level of < 10.0 g/dl 2.3% (1.6% to 3.1%), positive FOBT 7.1% (5.1% to 10%) and a blood glucose concentration of 410 mmol/l 0.78% (0.51% to 1.1%); all *p*-values < 0.001. See figure 2 of the publication[81] for additional PPVs

TABLE 39 Systematic review 2: included studies, model development and performance (*continued*)

Study	Model development	Model performance
Elias 2017[50]	N/A. Used RAT (colorectal) model[81]	• NPV: 99 (95% CI 98 to 100) • PPV: 8 (95% CI 6 to 10) • Sensitivity: 95 (95% CI 82 to 99) • Specificity: 45 (95% CI 41 to 49) • AUC: 0.81 (95% CI 0.75 to 0.88)
Hamilton 2009[74]	Variables with a univariable association with cancer significant with a *p*-value of < 0.1 were entered into a staged multivariable analysis	PPVs were estimated Six symptoms and two abnormal investigations (anaemia and microcytosis) were independently associated with CRC. The PPVs (95% CIs) of symptoms were as follows: rectal bleeding, to PPV for a male aged ≥ 80 years 4.5% (3.5% to 5.9%), change in bowel habit 3.9% (2.8% to 5.5%), weight loss 0.8% (0.5% to 1.3%), abdominal pain 1.2% (1.0% to 1.4%), diarrhoea 1.2% (1.0% to 1.5%) and constipation 0.7% (0.6% to 0.8%). PPVs were lower in females and younger patients. Only 27% of patients had reported either of the two higher-risk symptoms
Stapley 2017[72]	Conditional logistic regression. Variables independently associated with pancreatic cancer with a *p*-value of < 0.1 were entered into the multivariable analysis multivariable analysis performed in three stages; final model used a threshold *p*-value of < 0.02	PPVs were derived, using national incidence data to estimate prior odds (see figure 2 in publication[72])
Gastro-oesophageal		
Hippisley-Cox 2011[82]	Cox's proportional hazards. Rubin's rules were used to combine the results across the imputed data sets. Fractional polynomials were used to model non-linear risk relationships with continuous variables. Fitted full model, variables retained if HR < 0.80, HR > 1.20 and *p* < 0.01	Females: • R^2 (%): 71.2 (95% CI 69.2 to 73.2) • *D*-statistic: 3.22 (95% CI 3.06 to 3.37) • AUC: 0.89 (95% CI 0.87 to 0.91) Males: • R^2 (%): 72.5 (95% CI 71.1 to 73.8) • *D*-statistic: 3.32 (95% CI 3.21 to 3.43) • AUC: 0.92 (95% CI 0.91 to 0.93)
Collins 2013[88]	N/A. Used QCancer (gastro-oesophageal), see Hippisley-Cox 2011[82]	Men: Multiple imputation (*m* = 10) (*n* = 1,077,977) • R^2 (%): 74.4 (95% CI 73.0 to 75.8) • *D*-statistic: 3.49 (95% CI 3.35 to 3.62) • *c*-statistic: 0.93 (95% CI 0.92 to 0.94) Complete case (*n* = 852,532) • R^2 (%): 74.9 (95% CI 73.3 to 76.5) • *D*-statistic: 3.53 (95% CI 3.38 to 3.69) • *c*-statistic: 0.94 (95% CI 0.93 to 0.95) Women: Multiple imputation (*m* = 10) (*n* = 1,062,217) • R^2 (%): 75.6 (95% CI 74.6 to 76.5) • *D*-statistic: 3.60 (95% CI 3.51 to 3.69) • *c*-statistic: 0.94 (95% CI 0.94 to 0.95)

continued

TABLE 39 Systematic review 2: included studies, model development and performance (*continued*)

Study	Model development	Model performance
		Complete case ($n = 1,075,775$) • R^2 (%): 65.81 (95% CI 64.62 to 67.01) • *D*-statistic: 2.84 (95% CI 2.76 to 2.92) • *c*-statistic: 0.909 (95% CI 0.903 to 0.915)
Stapley 2013[71]	Conditional logistic regression. Variables independently associated with oesophago-gastric cancer with a *p*-value of < 0.1 were entered into the multivariable analysis. Significant variables (at $p < 0.05$) were grouped together and those variables significant at a *p*-value of < 0.01 remained in the final model	PPVs for the risk of oesophago-gastric cancer in patients consulting in primary care. PPVs (%): dysphagia 1.5 (95% CI 1.3 to 1.8), dyspepsia 0.2 (95% CI 0.19 to 0.22), nausea/vomiting 0.18 (0.17 to 0.20), abdominal pain 0.08 (95% CI 0.08 to 0.09), reflux 0.19 (95% CI 0.18 to 0.22), chest pain 0.05 (95% CI 0.05 to 0.06), epigastric pain 0.28 (95% CI 0.24 to 0.32), weight loss 0.26 (95% CI 0.23 to 0.31), constipation 0.07 (95% CI 0.06 to 0.07), low Hb 0.07 (95% CI 0.07 to 0.08), abnormal hepatic enzymes 0.04 (95% CI 0.04 to 0.05), raised inflammatory markers 0.08 (95% CI 0.08 to 0.09) and raised cholesterol 0.02 (95% CI 0.02 to 0.02)
Lung		
Iyen-Omofoman 2013[21]	Analyses conducted separately for features 4–12 and 13–24 months before diagnosis/index date Logistic regression. Included features statistically significant ($p < 0.05$) in univariate analyses. From multivariate model removed variables that were not significant, and features not significant in univariate analyses were rechecked for significance in multivariate model	Model based on features 4–12 months before diagnosis/index date. Area under the ROC curve of 0.88 Sensitivities (%) and specificities (%), respectively, reported at different risk score cut-off values: • −3: 93.98, 59.67 • −2.5: 88.31, 70.43 • −2: 79.57, 78.81 • −1.5: 68.40, 86.05 • −1.25: 61.52, 89.51 • −1: 53.07, 92.13 • −0.5: 35.30, 96.04 • 0: 21.24, 98.32 • 0.5: 10.07, 99.36
Hippisley-Cox 2011[52]	Cox proportional hazards models with age as the underlying time variable were used to develop separate risk equations in males and females Variables included if HR < 0.80, HR > 1.20 and a *p*-value of 0.01	Females: • R^2 (%): 71.70 (95% CI 70.30 to 73.10) • *D*-statistic: 3.25 (95% CI 3.15 to 3.37) • AUC: 0.92 (95% CI 0.91 to 0.93) Males: • R^2 (%): 72.11 (95% CI 71.04 to 73.18) • *D*-statistic: 3.29 (95% CI 3.20 to 3.38) • AUC: 0.92 (95% CI 0.91 to 0.93)
Hamilton 2005[51]	Variables associated with cancer in univariable analyses, using a *p*-value of ≤ 0.1, entered the multivariable analysis 18 clinically plausible interactions were tested in the final model; analyses were repeated excluding data from the last 180 days of the 730-day period studied. PPVs for individual variables and for pairs of variables were calculated from the LR and the observed incidence of cancer during the study	Colour-coded PPVs are reported in figure 2 of the publication[51]

TABLE 39 Systematic review 2: included studies, model development and performance (*continued*)

Study	Model development	Model performance
Ovarian		
Hippisley-Cox 2011[84]	Cox's proportional hazards model was used to estimate the coefficients for each risk factor using robust variance estimates to allow for the clustering of patients within general practices. Rubin's rules were used to combine the results across the imputed data sets. Fractional polynomials for modelling non-linear risk relations with continuous variables. Fitted full model, variables retained if HR < 0.80, HR > 1.20 and a *p*-value of 0.01	• R^2 (%): 57.6 (95% CI 54.8 to 60.4) • *D*-statistic: 2.38 (95% CI 2.24 to 2.51) • AUC: 0.84 (95% CI 0.83 to 0.86)
Collins 2013[89]	N/A. Used QCancer (ovarian), see Hippisley-Cox 2011[84]	• R^2 (%): 59.9 (95% CI 57.7 to 62.0) • *D*-statistic: 2.50 (95% CI 2.38 to 2.62) • AUC: 0.86 (95% CI 0.84 to 0.87)
Hamilton 2009[83]	Symptoms independently associated with the outcome, with a *p*-value of < 0.1, were retained for multivariable analyses. The resulting 99 variables were placed in eight groups and each group was analysed by multivariable conditional logistic regression. Symptoms still associated with cancer were rearranged into two larger groups (abdominal symptoms and other symptoms) for the final model. Discarded symptoms were checked against the final model. Five clinically plausible interactions were tested in the final model	PPVs were estimated and presented in figure 2 of the publication[83]
Pancreas		
Hippisley-Cox 2012[32]	Cox proportional hazards models with age as the underlying time variable were used to develop separate risk equations in males and females Variables included if HR < 0.80, HR > 1.20 and *p* < 0.01	Females: • R^2 (%): 58.7 (95% CI 55.4 to 61.9) • *D*-statistic: 2.44 (95% CI 2.27 to 2.60) • AUC: 0.84 (95% CI 0.82 to 0.86) Males: • R^2 (%): 62.0 (95% CI 59.1 to 64.8) • *D*-statistic: 2.61 (95% CI 2.45 to 2.77) • AUC: 0.87 (95% CI 0.85 to 0.88)

Collins 2013[14]	N/A. Used QCancer (pancreas), see Hippisley-Cox 2012[32]	Multiple imputation ($m = 10$)
		• Women ($n = 1{,}082{,}730$)
		  ○ R^2 (%): 60.0 (95% CI 56.6 to 63.5)
		  ○ *D*-statistic: 2.51 (95% CI 2.32 to 2.70)
		  ○ *c*-statistic: 0.89 (95% CI 0.87 to 0.90)
		• Men ($n = 1{,}067{,}592$)
		  ○ R^2 (%): 66.6 (95% CI 64.1 to 69.2)
		  ○ *D*-statistic: 2.89 (95% CI 2.72 to 3.07)
		  ○ *c*-statistic: 0.92 (95% CI 0.91 to 0.93)
		Complete case
		• Women ($n = 873{,}026$)
		  ○ R^2 (%): 62.0 (95% CI 58.4 to 65.7)
		  ○ *D*-statistic: 2.61 (95% CI 2.40 to 28.3)
		  ○ *c*-statistic: 0.90 (95% CI 0.88 to 0.91)

continued

TABLE 39 Systematic review 2: included studies, model development and performance (*continued*)

Study	Model development	Model performance

		• Men ($n = 823,873$)
		  ○ R^2 (%): 66.1 (95% CI 63.2 to 69.0)
		  ○ *D*-statistic: 2.86 (95% CI 2.66 to 3.05)
		  ○ *c*-statistic: 0.91 (95% CI 0.90 to 0.93)

Stapley 2012[70]	Conditional logistic regression. Variables independently associated with pancreatic cancer with a *p*-value of < 0.1 were entered into the multivariable analysis. In the first stage, similar variables were grouped together; in the second stage analyses were repeated with the new groups (with $p < 0.05$); in the third stage, the final model was built with the retained variables (with $p < 0.01$)	PPVs for the risk of pancreatic cancer in patients consulting in primary care were calculated using Bayes' theorem; see figure 2 in publication[70] PPVs for patients aged > 60 years were < 1%, apart from jaundice at 22% (95% CI 14% to 52%), although several pairs of symptoms had PPVs of > 1%
Keane 2014[62]	• Multivariable logistic regression was used to estimate ORs. Linear regression was used to estimate adjusted mean differences in clinical measures between patients with and patients without cancer • Ignored skewed laboratory test data on basis that sample size was large. Accounted for data clustering within practice	NR
Prostate		
Hamilton 2006[90]	Conditional logistic regression. Variables associated with cancer in univariable analyses, using a *p*-value of ≤ 0.1, entered the multivariable analysis. In the first stage, similar variables were grouped together; in the second stage, analyses were repeated with the new groups (with $p < 0.05$); in the third stage, the final model was built with the retained variables (with $p < 0.01$)	PPVs were calculated from the LRs and the observed annual incidence of cancer during the study, see figure 2 in publication[90] Eight features were associated with prostate cancer before diagnosis. Their PPVs against a background risk of 0.35% were as follows: urinary retention 3.1% (95% CI 1.5% to 6.0%); impotence 3.0% (95% CI 1.7% to 4.9%); frequency 2.2% (95% CI 1.3% to 3.5%); hesitancy 3.0% (95% CI 1.5% to 5.5%); nocturia 2.2% (95% CI 1.2% to 3.6%); haematuria 1.0% (95% CI 0.57% to 1.8%); weight loss 0.75% (95% CI 0.38% to 1.4%); abnormal rectal examination, deemed benign 2.8% (95% CI 1.6% to 4.6%); abnormal rectal examination, deemed malignant 12% (95% CI 5.0% to 37%). All $p < 0.001$, except for hesitancy ($p = 0.032$), nocturia ($p = 0.004$) and haematuria ($p = 0.009$)
Renal		
Hippisley-Cox 2012[85]	Cox proportional hazards models with age as the underlying time variable were used to develop separate risk equations in males and females Variables included if HR < 0.80, HR > 1.20 and a *p*-value of 0.01	Females: • R^2 (%): 74.8 (95% CI 73.2 to 76.5) • *D*-statistic: 3.53 (95% CI 3.37 to 3.68) • AUC: 0.91 (95% CI 0.90 to 0.93) Males: • R^2 (%): 75.5 (95% CI 74.6 to 76.4) • *D*-statistic: 3.60 (95% CI 3.51 to 3.69) • AUC: 0.95 (95% CI 0.94 to 0.96)

Collins 2013[91]	N/A. Used QCancer (renal), see Hippisley-Cox 2012[85]	Multiple imputation ($m = 10$)
		
		• Women:
		  ○ R^2 (%): 74.4 (95% CI 73.0 to 75.8)
		  ○ *D*-statistic: 3.49 (95% CI 3.36 to 3.62)
		  ○ *c*-statistic: 0.92 (95% CI 0.91 to 0.94)

TABLE 39 Systematic review 2: included studies, model development and performance (*continued*)

Study	Model development	Model performance

		• Men:
		  ○ R^2 (%): 74.2 (95% CI 73.4 to 75.1)
		  ○ *D*-statistic: 3.47 (95% CI 3.40 to 3.55)
		  ○ *c*-statistic: 0.95 (95% CI 0.94 to 0.95)
		Complete case
		• Women:
		  ○ R^2 (%): 74.0 (95% CI 72.1 to 75.5)
		  ○ *D*-statistic: 3.43 (95% CI 3.28 to 3.59)
		  ○ *c*-statistic: 0.92 (95% CI 0.91 to 0.94)
		• Men:
		  ○ R^2 (%): 74.0 (95% CI 73.0 to 75.1)
		  ○ *D*-statistic: 3.46 (95% CI 3.36 to 3.55)
		  ○ *c*-statistic: 0.95 (95% CI 0.94 to 0.96)

Shephard 2013[66]	Conditional logistic regression: features associated with cancer with *p*-values of ≤ 0.1 were grouped for the first multivariable analysis, with retention requiring a *p*-value of ≤ 0.05. A final multivariable model was compiled from the features of the previous stage for which $p < 0.01$	• PPVs were produced for all features shown to be independently associated with kidney cancer, see figure 2 of publication[66] • The PPV for visible haematuria (the most powerful single predictor) in patients aged ≥ 60 years was 1.0% (95% CI 0.8% to 1.3%)
Uterine		
Walker 2013[73]	Conditional logistic regression. Variables independently associated with pancreatic cancer with a *p*-value of < 0.1 were entered into the multivariable analysis. In the first stage, similar variables were grouped together; in the second stage, analyses were repeated with the new groups (with $p < 0.05$); in the third stage, the final model was built with the retained variables (with $p < 0.05$)	PPVs were estimated for features shown to be independently associated with uterine cancer in the multivariable analysis, using Bayes' theorem. See figure 1 in publication[73]
Metastatic		
Hamilton 2015[86]	Conditional logistic regression. Univariable analyses were performed initially, retaining variables with a *p*-value of < 0.1 to enter into multivariable analyses; only variables that were present in > 2% of the cases were studied. Cancer controls and healthy controls were used in separate analyses, and the cancer sites were analysed separately, and also merged and a unified analysis performed. Clinically plausible interaction terms were added to each model, and LR testing was applied to test whether or not they improved the models	NR
Multiple		
Hippisley-Cox 2013[53]	Multinomial logistic regression was used to estimate the coefficients for each predictor variable for each type of cancer. Fitted full model and variables retained if $p \leq 0.01$. Fractional polynomials were used to model non-linear risk relationships with continuous variables	• Discrimination (table 6): all ROC statistic values for each cancer type were > 0.79, except for cervix (0.73). The highest ROC values were for lung cancer (0.91) and uterine cancer (0.91) • Calibration (figure 1): compared mean predicted and observed risks. Overall, the model was well calibrated for each cancer type except for 'other cancer', which showed a degree of over prediction

continued

TABLE 39 Systematic review 2: included studies, model development and performance (*continued*)

Study	Model development	Model performance
		• Classification measures (table 7): symptoms with the highest PPVs for any cancer (regardless of type) were as follows: breast lump (11%), haemoptysis (8%), dysphagia (8%) and post-menopausal bleeding (7%). The PPV for anaemia was 6% and for venous thromboembolism was 5% • Used validation cohort to define thresholds for 1%, 5% and 10% risk. Calculated sensitivity, specificity, NPV and PPV at these thresholds
Hippisley-Cox 2013[54]	Multinomial logistic regression was used to estimate the coefficients for each predictor variable for each type of cancer. Fitted full model and variables retained if $p \leq 0.01$. Fractional polynomials were used to model non-linear risk relationships with continuous variables	• Discrimination (table 6): all ROC statistic values for each cancer type were > 0.82, indicating very good discrimination. The highest ROC values were for renal tract cancer (0.94), gastro-oesophageal cancer (0.93) and lung cancer (0.92). The lowest was for testicular cancer (0.82) • Calibration (figure 1): compared mean predicted and observed risks. Overall, the model was well calibrated for each cancer type except for the 'other cancer' model, which showed a degree of over prediction • Classification measures (table 7): symptoms with the highest PPVs for any cancer (regardless of type) were anaemia (19%), urinary retention (14%), dysphagia (13%), haematuria (13%), weight loss (11%), neck lump (10%) and haemoptysis (10%). The PPV for venous thromboembolism was 6%. The sensitivity of single symptoms was generally low, with the highest value being 16% for abdominal pain • Used validation cohort to define thresholds for 1%, 5% and 10% risk. Calculated sensitivity, specificity, NPV and PPV at these thresholds
Dommett 2013[64]	Conditional logistic regression. Variables occurring in at least 2% of either cases or controls with a univariable *p*-value of ≤ 0.1 entered the multivariable conditional logistic regression. A *p*-value of < 0.01 was used for retention in the final model. PPVs were calculated using Bayes' theorem	12 predictors had a PPV of ≥ 0.04%
Muris 1995[56]	Variables for which $p < 0.25$ in univariate analyses were entered into multiple stepwise forward logistic regressions	NR
Elias 2017[50]	N/A. Used Muris 1996[56] model	• NPV: 99 (95% CI 97 to 100) • PPV: 6 (95% CI 4 to 8) • Sensitivity: 97 (95% CI 86 to 100) • Specificity: 28 (95% CI 25 to 31) • AUC: 0.62 (95% CI 0.54 to 0.70)

AUC, area under the curve; CART, classification and regression tree; HR, hazard ratio; LR, likelihood ratio; N/A, not applicable; NPV, negative predictive value; NR, not reported; OR, odds ratio; RF, random forest; ROC, receiver operating characteristic; SE, standard error; SVM, support vector machine.

TABLE 40 Systematic review 2: included studies, results

Study	Model evaluation	Results	Interpretation and discussion
Bladder			
Shephard 2012[69]	Development data set only	Symptoms: • Visible haematuria – LR 59 (51 to 67), OR 34 (29 to 41) • Dysuria – LR 9.4 (8.0 to 11), OR 4.1 (3.4 to 5.0) • Abdominal pain – LR 2.0 (1.8 to 2.3), OR 2.0 (1.6 to 2.4) • Constipation – LR 1.8 (1.6 to 2.0), OR 1.5 (1.2 to 1.9) Disease: • Urinary tract infection – LR 5.2 (4.8 to 5.8), OR 2.2 (2.0 to 2.5) Investigations: • Raised levels of creatinine – LR 1.8 (1.6 to 1.9), OR 1.3 (1.2 to 1.4) • Raised levels of inflammatory markers – LR 1.8 (1.6 to 2.1), OR 1.5 (1.2 to 1.9) • Raised WBC count – LR 2.8 (2.4 to 3.2), OR 2.1 (1.6 to 2.8)	Findings support investigation of all patients aged > 40 years with visible haematuria
Blood			
Shephard 2016[67]	Development data set only	Ten symptoms were independently associated with chronic leukaemia, OR (95% CIs): infection 1.5 (1.3 to 1.6), cough 1.2 (1.1 to 1.4), hypertension 1.2 (1.1 to 1.4), shortness of breath 1.3 (1.1 to 1.5), fatigue 2.1 (1.8 to 2.6), diarrhoea 1.4 (1.1 to 1.7), lymphadenopathy 22 (13 to 36), malaise 1.7 (1.3 to 2.3), weight loss 3.0 (2.1 to 4.2) and bruising 2.3 (1.6 to 3.2) Thirteen symptoms were independently associated with acute leukaemia, OR (95% CIs): infection 1.5 (1.3 to 1.8), shortness of breath 2.5 (1.9 to 3.2), fatigue 4.4 (3.3 to 6.0), chest pain 1.5 (1.1 to 2.1), abdominal pain 1.7 (1.2 to 2.2), diarrhoea 2.2 (1.5 to 3.1), malaise 3.4 (2.2 to 5.2), vomiting/nausea 1.8 (1.2 to 2.6), bruising 3.7 (2.3 to 5.8), fever 5.3 (2.7 to 10), nosebleeds and/or bleeding gums 5.7 (3.1 to 10), flu 3.9 (2 to 7.5) and weight loss 3 (1.5 to 5.8) No individual symptom or combination of symptoms had a PPV of > 1%	PPVs too small, with benign alternative explanations much more likely for the symptoms

continued

TABLE 40 Systematic review 2: included studies, results (*continued*)

Study	Model evaluation	Results	Interpretation and discussion
Shephard 2015[68]	Development data set only	Symptoms: • Back pain – LR4.6 (4.2 to 5.0), OR 2.2 (2.0 to 2.4) • Chest pain – LR 3.4 (3.0 to 3.8), OR 1.6 (1.4 to 1.8) • Chest infection – LR 1.9 (1.7 to 2.1), OR 1.4 (1.2 to 1.6) • Shortness of breath – LR 1.9 (1.7 to 2.2), OR 1.3 (1.1 to 1.5) • Nausea – LR 3.2 (2.6 to 3.9), OR 1.5 (1.1 to 2.1) • Fracture – LR 3.6 (2.9 to 4.6), OR 3.1 (2.3 to 4.2) • Joint pain – LR 1.5 (1.2 to 1.8), OR 1.6 (1.2 to 2.2) • Combined bone pain – LR 4.3 (3.3 to 5.6), OR 2.1 (1.4 to 3.1) • Weight loss – LR 5.6 (4.2 to 7.1), OR 3.0 (2.0 to 4.5) • Rib pain – LR 7.7 (5.4 to 11.0), OR 2.5 (1.5 to 4.4) • Nosebleeds – LR 4.4 (3.2 to 6.0), OR 3.0 (1.9 to 4.7) Investigations: • Cytopenia – LR 5.3 (5.0 to 5.7), OR 5.4 (4.6 to 6.4) • Raised levels of inflammatory markers – LR 6.8 (6.3 to 7.4), OR 4.9 (4.2 to 5.8) • Raised levels of creatinine – LR 2.9 (2.6 to 3.1), OR 1.8 (1.5 to 2.2) • Raised mean corpuscular volume – LR 6.2 (5.3 to 7.3), OR 3.1 (2.4 to 4.1) • Hypercalcaemia – LR 26 (18 to 35), OR 11.4 (7.1 to 18)	Although no single symptom is a strong indicator of myeloma, repeated occurrences of back pain or back pain combined with nosebleeds or rib pain suggest initial testing of inflammatory markers, at the discretion of the GP
Brain			
Hamilton 2007[87]	Development data set only	New-onset seizure 87.0 (42.0 to 180.0), weakness 23.0 (7.1 to 77.0), headache 6.7 (5.6 to 8.0), confusion 11.0 (7.6 to 16.0), memory loss 2.7 (1.7 to 4.2), visual disorder 2.0 (1.2 to 3.3), motor loss 1.8 (1.5 to 2.2) and motor loss with weakness 0.2 (0.06 to 0.8)	Findings suggest that isolated headache presented to primary care has too small a risk of an underlying brain tumour to warrant investigation; however, new-onset seizures should be investigated
Breast			
McCowan 2011[60]	External validation	Independent clinical predictors (adjusted OR and 95% CIs): increasing age by year, 1.10 (1.07 to 1.13); presence of a discrete lump, 15.20 (4.88 to 47.34); breast thickening, 7.64 (2.23 to 26.11); lymphadenopathy, 3.63 (1.33 to 9.92); and lump of 2 cm, 5.41 (2.36 to 12.38). All eight patients with skin tethering had breast cancer	The clinical prediction rule discriminates between patients at high risk and patients at low risk of breast cancer

TABLE 40 Systematic review 2: included studies, results (*continued*)

Study	Model evaluation	Results	Interpretation and discussion
Colorectal			
Marshall 2011[22]	• Apparent performance with THIN data set • BB validated with CAPER data set	Model predictors and ORs (95% CIs): constipation, 2.06 (1.88 to 2.26); diarrhoea, 2.38 (2.14 to 2.66); change in bowel habit, 13.83 (11.70 to 16.34); abdominal pain, 3.82 (3.49 to 4.18); rectal bleeding, 20.11 (17.35 to 23.32); Hb level of 13–13.99 g/dl, 1.33 (1.18 to 1.50); Hb level of 12–12.99 g/dl, 1.63 (1.42 to 1.87); Hb level of 11–11.99 g/dl, 2.54 (2.16 to 2.99); Hb level of 10–10.99 g/dl, 5.18 (4.19 to 6.39); Hb level of 9–9.99 g/dl, 8.08 (6.13 to 10.65); Hb level of < 9 g/dl, 15.94 (11.78 to 21.57); mean cell volume of 80–84.99 fl, 2.71 (2.30 to 3.19); mean cell volume of < 80 fl, 7.67 (6.23 to 9.44); weight loss of ≥ 10%, 2.92 (2.39 to 3.57); and weight loss of 5–10%, 1.37 (1.09 to 1.73)	Both multivariable BB and CAPER equations performed significantly better than NICE referral guidelines[172]
Elias 2017[50]	External validation	Elias ranked the BB equation as sixth out of 19 models evaluated	The top-ranked model was the NICE guidelines. Note that Elias used a very broad definition of prediction model, which included guidelines and weighted scores
Fijten 1995[55]	Development data set only	• Diagnostic index = −6.7 + 2.1 [(age − 50)/10] + 2.3 if change in bowel habit + 2.1 if blood mixed with or on stool. (Hosmer–Lemeshow $p = 0.507$), log likelihood = −17.9) • Cut-off point for maximising sensitivity (100%) and specificity (90%) was 0.042 • Diagnostic index with presence of polyps: −4.8 + 1.4 [(age − 50)/10] + 1.9 if change in bowel habit + 2.1 if blood mixed with or on stool. (Hosmer–Lemeshow $p = 0.49$), log likelihood = −32.4) • Cut-off point for maximising sensitivity and specificity was 0.058	The combination of age, change in bowel habit and blood seen mixed with or on stool can serve as a useful diagnostic tool for the prediction of CRC (and overtly bleeding polyps)
Hodder 2005[77]	External validation	The Netherlands model had better discrimination than the Harvard model, but inferior discrimination compared with the weighted numerical score	
Elias 2017[50]	External validation	CEDAR data set included participants older than those in original Fijten 1995[55] derivation data set. Based on the results, Elias *et al.* ranked the Fijten 1995[55] model as 13th out of 19 models	The top-ranked model was NICE guidelines. Note that Elias 2017 used a very broad definition of prediction model, which included guidelines and weighted scores

continued

TABLE 40 Systematic review 2: included studies, results (*continued*)

Study	Model evaluation	Results	Interpretation and discussion
Kop 2015[61]	Five subsets were selected and four machine-learning algorithms were applied to them in a fivefold cross-validation fashion: 1. Non-temporal (941 non-temporal attributes + age/sex) 2. Temporal (*n* temporal patterns + age/sex) 3. All (941 + *n* + age/sex) 4. Knowledge driven (31 attributes as described in Marshall 2011[22] + age/sex) 5. Age/sex (benchmark; age and sex only)	AUCs and 95% CIs for subsets: • Non-temporal – LR 0.792 (0.771 to 0.813); RF 0.883 (0.866 to 0.900); SVM 0.804 (0.784 to 0.824); CART 0.819 (0.799 to 0.839) • Temporal – LR 0.893 (0.877 to 0.909); RF 0.882 (0.865 to 0.899); SVM 0.861 (0.843 to 0.879); CART 0.863 (0.845 to 0.881) • All – LR 0.796 (0.775 to 0.817); RF 0.881 (0.864 to 0.898); SVM 0.832 (0.813 to 0.851); CART 0.818 (0.798 to 0.838) • Knowledge driven – LR 0.854 (0.836 to 0.872); RF 0.896 (0.880 to 0.912); SVM 0.867 (0.849 to 0.885); CART 0.860 (0.842 to 0.878) • Age/sex only – LR 0.844 (0.825 to 0.863); RF 0.838 (0.819 to 0.857); SVM 0.862 (0.844 to 0.880); CART 0.828 (0.808 to 0.848)	Study suggests metabolic syndrome as potential predictor
Nørrelund 1996[57]	Validated in study 2	Only age was found to be a statistically significant predictor of cancer from study 1: • Age 70–79 years: adjusted OR 9.26 (95% CI 3.32 to 25.82); age ≥ 80 years: adjusted OR 9.90 (95% CI 2.03 to 48.36) No statistically significant variables were found in study 2 to predict cancer	Reported symptoms of weight loss, abdominal pain, change in bowel habits or discomfort were not found to be predictors of CRC in either study 1 or 2
Elias 2017[50]	External validation	Elias *et al.*[50] ranked the Danish (i.e. Nørrelund *et al.*[57]) model as 18th out of 19 models evaluated	The top-ranked model was NICE guidelines. Note that Elias 2017 used a very broad definition of prediction model, which included guidelines and weighted scores
Hippisley-Cox 2012[80]	Split sample validation (66%/33%)	Predictors and (fully adjusted) HR (95% CI): • Females – family history of GI cancer 1.39 (1.02 to 1.89), Hb level of < 11 g/dl 3.26 (2.84 to 3.74), current rectal bleeding 32.3 (27.7 to 37.6), current abdominal pain 6.90 (5.91 to 8.06), current appetite loss 2.43 (1.70 to 3.47), current weight loss 7.70 (5.32 to 11.1) • Males – trivial drinker 1.07 (0.95 to 1.20), light drinker 1.20 (1.06 to 1.35), moderate/heavy drinker 1.43 (1.25 to 1.63), family history of GI	State that the algorithm performed well, with good discrimination and calibration

TABLE 40 Systematic review 2: included studies, results (*continued*)

Study	Model evaluation	Results	Interpretation and discussion
		cancer 1.52 (1.12 to 2.07), Hb level of < 11 g/dl 3.33 (2.86 to 3.87), current rectal bleeding 27.0 (23.5 to 31.1), current abdominal pain 6.78 (5.76 to 7.97), current appetite loss 2.15 (1.53 to 3.03), current weight loss 4.07 (3.42 to 4.85), change in bowel habit in previous years 2.25 (1.47 to 3.46)	
Collins 2012[15]	External validation with a different cohort	Model calibration is very good, with close agreement between predicted and observed CRC risks across all tenths of risk	
Hamilton 2005[81]	Development data set only	ORs (95% CIs): rectal bleeding 15 (9.0 to 2.4), weight loss 2.7 (1.7 to 4.6), number of episodes of abdominal pain 2.2 (1.7 to 2.8), constipation 2.0 (1.2 to 3.3), number of episodes of diarrhoea 1.6 (1.3 to 2.0), rectal disease on examination 13 (4.7 to 37), tenderness on palpitation of abdomen 3.6 (1.7 to 7.8), positive FOBT 81 (20 to 330), Hb level of 12.0–12.9 g/dl 2.5 (0.95 to 6.8), Hb level of 10.0–11.9 g/dl 4.3 (2.1 to 9.0), Hb level of < 10 g/dl 13 (6.2 to 28), blood sugar concentration of > 10 mmol/l 2.0 (1.3 to 3.1) Interaction terms, ORs (95% CIs): abdominal pain with tenderness 0.56 (0.38 to 0.82), positive FOBT with a Hb level of < 10 g/dl 0.020 (0.0015 to 0.27)	10 symptoms, signs or investigation results were independently associated with CRC. Five of these remained associated with cancer 180 days before diagnosis
Elias 2017[50]	External validation	Elias ranked the RAT (CRC) model as 10th out of 19 models evaluated. The top-ranked model was the NICE guidelines	Note that Elias used a very broad definition of prediction model, which included guidelines and weighted scores
Hamilton 2009[74]	Development data set only	ORs (95% CIs): rectal bleeding 20 (17 to 23), change in bowel habit 14 (12 to 17), abdominal pain 3.9 (3.6 to 4.3), diarrhoea 2.4 (2.1 to 2.7), constipation 2.1 (1.9 to 2.3), weight loss of 5.0–9.9% 1.2 (0.99 to 1.5), weight loss of ≥ 10% 2.5 (2.1 to 3.0), Hb level of 12.0–12.9 g/dl 1.7 (1.5 to 1.9), Hb level of 11.0–11.9 g/dl 2.8 (2.4 to 3.2), Hb level of 10.0–10.9 g/dl 5.9 (4.8 to 7.2), Hb level of 9.0–9.9 g/dl 9.3 (7.1 to 12), Hb level of < 9 g/dl 18 (14 to 25), mean red cell volume of < 80 fl 6.5 (5.3 to 7.9)	There is a need to improve identification of CRC among the large number of patients presenting only with low-risk symptoms

continued

TABLE 40 Systematic review 2: included studies, results (*continued*)

Study	Model evaluation	Results	Interpretation and discussion
Stapley 2017[72]	Development data set only	OR (95% CIs) for CRC: diarrhoea 7.7 (4.3 to 14), abdominal pain 6.0 (4.2 to 8.7), rectal bleeding 54 (26 to 110), change in bowel habit 58 (21 to 160), constipation 7.9 (4.3 to 14), nausea/vomiting 2.7 (1.4 to 5.1), rectal mass 190 (51 to 720), raised levels of inflammatory markers 3.1 (2.0 to 4.7), low Hb level 5.2 (3.2 to 8.5), low mean red cell volume 4.3 (2.3 to 8.0)	*Rectal bleeding and change in bowel habit are strongly predictive of CRC/IBD when combined with abnormal haematology*
Gastro-oesophageal			
Hippisley-Cox 2011[82]	Split sample validation (66%/33%)	Predictors and (fully adjusted) HR (95% CI) • Females – ex-smoker 1.33 (1.11 to 1.60), light smoker 1.96 (1.43 to 2.68), moderate smoker 2.51 (1.93 to 3. 26), heavy smoker 3.11 (2.26 to 4.28), current dysphagia 131 (97.5 to 175), current abdominal pain 4.74 (3.54 to 6.33), current appetite loss 10 (5.28 to 19), current haematemesis 25.2 (14.4 to 44.2), current weight loss 3.97 (3.06 to 5.16), Hb level of < 11 g/dl 2.32 (1.84 to 2.93) • Males – ex-smoker 1.38 (1.22 to 1.57), light smoker 1.89 (1.51 to 2.37), moderate smoker 2.18 (1.77 to 2.67), heavy smoker 2.00 (1.52 to 2.63), current dysphagia 143 (108 to 189), current abdominal pain 378 (3.32 to 4.30), current appetite loss 3.87 (2.82 to 5.32), current haematemesis 7.62 (6.08 to 9.55), current weight loss 5.64 (4.67 to 6.81), Hb level of < 11 g/dl 1.79 (1.44 to 2.23)	The algorithm developed and validated in the study performed well in quantifying the absolute risk of having existing, but as yet undiagnosed, gastro-oesophageal cancer
Collins 2013[88]	External validation	Model calibration was good, with reasonable agreement between predicted and observed gastro-oesophageal cancer risks across all but the last tenth of risk	
Stapley 2013[71]	Development data set only	Sixteen features were independently associated with oesophagogastric cancer, OR (95% CI) (all $p < 0.001$): dysphagia, 139 (112 to 173); reflux, 5.7 (4.8 to 6.8); abdominal pain, 2.6 (2.3 to 3.0); epigastric pain, 8.8 (7.0 to 11.0); dyspepsia, 6 (5.1 to 7.1); nausea and/or vomiting, 4.9 (4.0 to 6.0); constipation, 1.5 (1.2 to 1.7); chest pain, 1.6 (1.4 to 1.9); weight loss, 8.9 (7.1 to 11.2);	The tool can be used to guide GPs on referral for oesophagogastric cancer. Evidence suggests that reliance on 'red-flag' symptoms solely has limited use

TABLE 40 Systematic review 2: included studies, results (*continued*)

Study	Model evaluation	Results	Interpretation and discussion
		thrombocytosis, 2.4 (2.0 to 2.9); low level of Hb, 2.4 (2.1 to 2.7); low MCV, 5.2 (4.2 to 6.4); high levels of inflammatory markers, 1.7 (1.4 to 2.0); raised levels of hepatic enzymes, 1.3 (1.2 to 1.5); high WBC count, 1.4 (1.2 to 1.7); and high cholesterol, 0.8 (0.7 to 0.8)	
Lung			
Iyen-Omofoman 2013[21]	Split sampling validation	Clinical and sociodemographic features that were independently associated with lung cancer were patients' age, sex, socioeconomic status and smoking history. From 4 to 12 months before diagnosis, the symptoms/clinical features that were independently predictive of lung cancer were as follows, OR (95% CI): 11–20 GP consultations 1.23 (1.16 to 1.29), > 20 GP consultations 1.36 (1.28 to 144), cough 1.63 (1.53 to 1.75), haemoptysis 8.70 (6.75 to 11.20), dyspnoea 1.41 (1.29 to 1.55), weight loss 2.66 (2.13 to 3.29), lower respiratory tract infections 1.56 (1.38 to 1.76), non-specific chest infections 1.55 (1.44 to 1.68), chest/shoulder pain 1.39 (1.28 to 1.51), hoarseness 1.79 (1.28 to 2.49), upper respiratory tract infections 1.15 (1.02 to 1.30) and chronic obstructive pulmonary disease 1.61 (1.46 to 1.78)	Risk prediction model performed better than existing NICE guidelines
Hippisley-Cox 2011[52]	Split sample validation (66%/33%)	Predictors and (fully adjusted) HR (95% CI): • Females – current haemoptysis 23.9 (20.6 to 27.6), current appetite loss 4.14 (3.15 to 5.45), current weight loss 4.52 (3.80 to 5.38), cough in previous 12 months 1.90 (1.56 to 2.32), Hb level of < 11 g/dl 1.75 (1.38 to 2.22), ex-smoker 3.37 (2.38 to 4.01), light smoker 6.57 (5.37 to 8.03), moderate smoker 8.32 (7.05 to 9.82), heavy smoker 10.6 (8.49 to 13.2), prior diagnosis non-lung cancer 1.33 (1.09 to 1.63), chronic obstructive airways disease 1.82 (1.57 to 2.11), Townsend deprivation score 1.17 (1.08 to 1.27)	The algorithm developed and validated in the study performed well in quantifying the absolute risk of having existing, but as yet undiagnosed, lung cancer

continued

TABLE 40 Systematic review 2: included studies, results (*continued*)

Study	Model evaluation	Results	Interpretation and discussion
		• Males – current haemoptysis 21.5 (19.3 to 23.9), current appetite loss 4.71 (3.69 to 6.00), current weight loss 6.09 (5.33 to 6.95), cough in previous 12 months 1.47 (1.23 to 1.75), Hb level of < 11 g/dl 1.89 (1.54 to 2.32), ex-smoker 2.13 (1.87 to 2.43), light smoker 3.70 (3.20 to 4.27), moderate smoker 4.95 (4.26 to 5.76), heavy smoker 6.35 (5.43 to 7.43), chronic obstructive airways disease 1.51 (1.34 to 1.69), Townsend deprivation score 1.17 (1.10 to 1.24)	
Hamilton 2005[51]	Development data set only	ORs (95% CIs): appetite loss 86 (3.6 to 2100), haemoptysis 32 (13 to 81), dyspnoea 4.7 (2.7 to 8.0), weight loss 4.3 (2.2 to 8.2), fatigue 3.2 (1.7 to 6.0), chest pain 2.9 (1.8 to 4.7), second attendance with cough 2.7 (1.7 to 4.4), finger clubbing 18 (1.7 to 190), thrombocytosis 9.3 (3.4 to 26), abnormal spirometry 7.5 (2.8 to 21), current smoker 9.7 (5.3 to 18), ex-smoker 5.9 (3.0 to 12), smoking status unknown 5.4 (2.8 to 10), dyspnoea with fatigue 0.28 (0.11 to 0.73), appetite loss in patients aged > 70 years 0.13 (0.024 to 0.76) After excluding variables reported in the final 180 days before diagnosis, haemoptysis, dyspnoea and abnormal spirometry remained independently associated with cancer	The study provides an evidence base for selection of patients for investigation of possible lung cancer, both for clinicians and for developers of guidelines
Ovarian			
Hippisley-Cox 2011[84]	Split sample validation (66%/33%)	Predictors and (fully adjusted) HRs (95% CIs): family history of ovarian cancer 9.8 (5.4 to 17.9), Hb level of < 110 g/l in the previous year 2.3 (1.7 to 2.9), current abdominal pain 7.0 (6.1 to 8.0), current abdominal distension 23.1 (18.2 to 29.4), current appetite loss 5.2 (3.4 to 7.9), current rectal bleeding 2.0 (1.4 to 2.8), current postmenopausal bleeding 6.6 (5.1 to 8.5), current weight loss 2.0 (1.3 to 3.1)	The algorithm developed and validated in the study performed well in quantifying the absolute risk of having existing, but as yet undiagnosed, ovarian cancer
Collins 2013[89]	External validation	Model calibration was good with reasonable agreement between predicted and observed ovarian cancer risks across all tenths of risk, with a slight overprediction Good predictive ability for identifying patients with undetected ovarian cancer	

TABLE 40 Systematic review 2: included studies, results (*continued*)

Study	Model evaluation	Results	Interpretation and discussion
Hamilton 2009[83]	Development data set only	Seven symptoms and one interaction term were associated with ovarian cancer in multivariable analysis. The univariable PPVs (95% CIs) and multivariable ORs (95% CIs), respectively, for these were 2.5% (1.2% to 5.9%) and 240 (46 to 1200) for abdominal distension; 0.5% (0.2% to 0.9%) and 24 (9.3 to 64) for postmenopausal bleeding; 0.6% (0.3% to 1.0%) and 17 (6.1 to 50) for loss of appetite; 0.2% (0.1% to 0.3%) and 16 (5.6 to 48) for increased urinary frequency; 0.3% (0.2% to 0.3%) and 12 (6.1 to 22) for abdominal pain; 0.2% (0.1% to 0.4%) and 7.6 (2.5 to 23) for rectal bleeding; and 0.3% (0.2% to 0.6%) and 5.3 (1.8 to 16) for abdominal bloating Interaction term OR (95% CI): abdominal distension with urinary frequency 0.015 (0 to 0.29) In 181 (85%) cases and 164 (15%) controls, at least one of these seven symptoms was reported to primary care before diagnosis. After exclusion of symptoms reported in the 180 days before diagnosis, abdominal distension, urinary frequency and abdominal pain remained independently associated with a diagnosis of ovarian cancer	Study findings suggest that early symptoms may be useful in the identification of ovarian cancer
Pancreas			
Hippisley-Cox 2012[32]	Split sample validation (66%/33%)	Predictors and (fully adjusted) HR (95% CI): • Females – ex-smoker 0.97 (0.77 to 1.23), light smoker 1.53 (1.04 to 2.25), moderate smoker 2.32 (1.74 to 3.10), heavy smoker 2.39 (1.65 to 3.48), type 2 diabetes 2.07 (1.66 to 2.58), chronic pancreatitis 3.15 (1.17 to 8.46), current appetite loss 3.90 (2.61 to 5.82), current weight loss 3.27 (2.35 to 4.56), current abdominal pain 4.09 (3.46 to 4.84), current abdominal distension 3.04 (1.68 to 5.50) • Males – ex-smoker 1.37 (1.12 to 1.67), light smoker 1.44 (1.03 to 2.03), moderate smoker 1.63 (1.20 to 2.20),	The algorithm developed and validated in the study performed well in quantifying the absolute risk of having existing, but as yet undiagnosed, pancreatic cancer

continued

TABLE 40 Systematic review 2: included studies, results (*continued*)

Study	Model evaluation	Results	Interpretation and discussion
		heavy smoker 1.88 (1.36 to 2.61), type 2 diabetes 2.11 (1.76 to 2.52), chronic pancreatitis 3.94 (1.93 to 8.01), current appetite loss 2.46 (1.43 to 4.23), current weight loss 12.5 (7.84 to 19.9), current abdominal pain 5.23 (4.48 to 6.11), current dysphagia 2.56 (1.60 to 4.10), constipation in previous year 1.91 (1.35 to 2.71)	
Collins 2013[14]	External validation	QCancer (pancreas) increasingly overpredicts risk, with increases across the tenths of risk	
Stapley 2012[70]	Development data set only	Nine features were associated with pancreatic cancer, OR (95% CI) [all $p < 0.001$ except for back pain, ($p = 0.004$)]; jaundice, OR 1000 (95% CI 430 to 2500); abdominal pain, 5 (4.4 to 5.6); nausea/vomiting, 4.5 (3.5 to 5.7); back pain, 1.4 (1.1 to 1.7); constipation, 2.2 (1.7 to 2.8); diarrhoea, 1.9 (1.5 to 2.5); weight loss, 15 (11 to 22); malaise, 2.4 (1.6 to 3.5); and new-onset diabetes 2.1 (1.7 to 2.5)	Most previously reported symptoms of pancreatic cancer were also relevant in primary care, providing a basis for selection of patients for investigation, especially with multiple symptoms, although predictive values were mostly small
Keane 2014[62]	Development data set only	• Independent predictors and adjusted (age, sex, time, deprivation) PDAC symptoms and signs, OR (95% CIs): weight loss 6.6 (5.54 to 7.86); abdominal pain 6.38 (5.81 to 7.02); nausea and vomiting 3.43 (3.00 to 3.91); bloating 3.1 (2.48 to 3.89); dyspepsia 2.56 (2.30 to 2.85); new-onset diabetes 2.46 (2.16 to 2.80); change in bowel habit 2.17 (1.98 to 2.39); pruritus 1.73 (1.43 to 2.10); lethargy 1.42 (1.25 to 1.61); back pain 1.33 (1.18 to 1.49); shoulder pain 0.78 (0.65 to 0.93); and jaundice 246 (172 to 351) • PDAC tests and BMI coefficient: bilirubin 15.3 (95% CI 12.99 to 17.60) • BTC symptoms and signs, OR (95% CIs): weight loss 3.17 (2.32 to 4.34); abdominal pain 4.68 (4.01 to 5.47); nausea and vomiting 2.99 (2.44 to 3.66); bloating 2.35 (1.57 to 3.53); dyspepsia 21.70 (1.40 to 2.08); change in bowel habit 1.77 (1.51 to 2.09); pruritus 3.75 (2.96 to 4.74); and jaundice 445 (302 to 658) • Bilirubin 15.3 (95% CI 12.99 to 17.60)	Findings could improve the symptom-based cancer decision support tools used for identification of BTC and PDAC

TABLE 40 Systematic review 2: included studies, results (*continued*)

Study	Model evaluation	Results	Interpretation and discussion
Prostate			
Hamilton 2006[90]	Development data set only	ORs (95% CIs): urinary retention 11 (5 to 25.5), second presentation with weight loss 9.2 (2.7 to 31), impotence 5.3 (2.8 to 9.8), frequency 3.2 (1.9 to 5.4), hesitancy 2.9 (1.1 to 7.5), nocturia 2.6 (1.3 to 5), haematuria 2.4 (1.3 to 4.7), benign abnormal rectal examination 3.7 (1.9 to 7.3), malignant abnormal rectal examination 70 (13 to 380) Loss of weight, impotence, frequency and abnormal rectal examination remained associated with cancer after excluding the final 180 days from analysis	The predictive values for symptoms present in most men with prostate cancer will help guide GPs and patients about the value of further investigation
Renal			
Hippisley-Cox 2012[85]	Split sample validation (66%/33%)	Predictors and (fully adjusted) HR (95% CI): • Females – ex-smoker 1.23 (1.01 to 1.49), light smoker 1.83 (1.31 to 2.57), moderate smoker 2.41 (1.87 to 3.12), heavy smoker 2.32 (1.55 to 3.46), history of prior cancer (not renal) 1.47 (1.13 to 1.91), current haematuria 119 (85.3 to 167), current appetite loss 2.45 (1.34 to 4.46), current abdominal pain 2.38 (1.97 to 2.89), current weight loss 2.56 (1.75 to 3.74), anaemia 1.98 (1.51 to 2.61) • Males – ex-smoker 1.47 (1.32 to 1.63), light smoker 2.24 (1.84 to 2.73), moderate smoker 2.49 (2.05 to 3.04), heavy smoker 2.50 (1.95 to 3.20), current haematuria 148 (123 to 177), current abdominal pain 3.06 (2.35 to 3.98), current weight loss 5.67 (3.20 to 10.0), anaemia 1.57 (1.27 to 1.94)	The algorithm developed and validated in the study performed well in quantifying the absolute risk of having existing, but as yet undiagnosed, renal cancer
Collins 2013[91]	External validation	Model calibration showed good agreement across those aged 30–69 years, whereas it overpredicted the risk of renal tract cancer in those aged between 70 and 84 years The model showed good predictive ability for identifying patients with suspected undiagnosed renal tract cancer who would benefit from further clinical investigation	

continued

TABLE 40 Systematic review 2: included studies, results (*continued*)

Study	Model evaluation	Results	Interpretation and discussion
Shephard 2013[66]	Development data set only	Fifteen features were independently associated with kidney cancer, OR (95% CI): visible haematuria 37 (28 to 49), abdominal pain 2.8 (2.4 to 3.4), microcytosis 2.6 (1.9 to 3.4), raised inflammatory markers 2.4 (2.1 to 2.8), thrombocytosis 2.2 (1.7 to 2.7), low level of Hb 1.9 (1.6 to 2.2), urinary tract infection 1.8 (1.5 to 2.1), nausea 1.8 (1.4 to 2.3), raised creatinine levels 1.7 (1.5 to 2.0), leukocytosis 1.5 (1.2 to 1.9), fatigue 1.5 (1.2 to 1.9), constipation 1.4 (1.1 to 1.7), back pain 1.4 (1.2 to 1.7), abnormal liver function 1.3 (1.2 to 1.5) and raised blood sugar 1.2 (1.1 to 1.4)	Visible haematuria is the commonest and most powerful single predictor of kidney cancer, and the risk rises when additional symptoms are present
Uterine			
Walker 2013[73]	Development data set only	Nine features were significantly associated with uterine cancer, OR (95% CI): postmenopausal bleeding 154.5 (100.4 to 237.8), excessive vaginal bleeding 22.3 (11.9 to 41.8), irregular menstruation (one GP visit) 41.5 (27.3 to 63.0), irregular menstruation (two GP visits) 69.0 (31.6 to 150.7), vaginal discharge 13.7 (10 to 21), haematuria 8.7 (5.0 to 15.1), abdominal pain (one GP visit) 2.0 (1.4 to 2.8), abdominal pain (two GP visits) 3.4 (2.0 to 5.8), low Hb 2.1 (1.5 to 2.9), raised platelets 1.5 (1.0 to 2.3) and raised glucose 1.4 (1.1 to 1.8); all $p < 0.01$, other than raised platelet count ($p = 0.05$) and raised glucose concentrations ($p = 0.02$) In the year before diagnosis, 1725 (63%) cases had a record of abnormal vaginal bleeding, compared with 135 (1%) controls. The PPV of uterine cancer with postmenopausal bleeding was 4%, and was higher in women with multiple or repeated symptoms	Findings of the study, in particular the importance of postmenopausal bleeding and haematuria, for uterine cancer may inform GPs in the selection of women for investigation
Metastatic			
Hamilton 2015[86]	Development data set only	Adjusted OR (95% CI): • Vs. cancer controls: groin pain 10.2 (1.2 to 8.2), pleurisy/ pleural effusion 10.2 (1.1 to 9.2), shoulder pain 5.3 (1.6 to 1.8), loss of appetite 4.0 (1.2 to 1.3), vomiting 3.5 (1.3 to 9.4), low-back pain 2.5 (1.1 to 5.6), abnormal liver function 3.5 (1.6 to 7.5)	The scarcity of specific symptoms and the fairly common occurrence of non-specific symptoms (vomiting and loss of appetite) may explain delays in the diagnosis of metastases

TABLE 40 Systematic review 2: included studies, results (*continued*)

Study	Model evaluation	Results	Interpretation and discussion
		• Vs. healthy controls: vomiting 3.6 (1.3 to 1.0), 4.2 (1.5 to 1.2), flank/loin pain 19.4 (1.8 to 2.10), chest pain musculoskeletal 5.3 (1.7 to 1.6), oedema 3.4 (1.1 to 10), abnormal liver function 5.1 (1.9 to 1.4)	
Multiple			
Hippisley-Cox 2013[53]	Split sample validation (66%/33%)	Each cancer model contained the following number of predictors (see additional tables on www.qcancer.org for details) for these cancers: breast 10, cervical 12, ovarian 11, uterine 7, blood 10, colorectal 13, gastro-oesophageal 13, lung 15, pancreatic 14, renal 11, other cancers 22	A new algorithm designed to estimate the absolute risk of having existing but as yet undiagnosed cancer was developed and validated
Hippisley-Cox 2013[54]	Split sample validation (66%/33%)	Each cancer model contained the following number of predictors (see additional tables on www.qcancer.org for details) for these cancers: prostate 14, testicular 3, blood 14, colorectal 12, gastro-oesophageal 13, lung 17, pancreatic 15, renal tract 8, other cancers 20	A new algorithm designed to estimate the absolute risk of having existing but as yet undiagnosed cancer was developed and validated
Dommett 2013[64]	Development data set only	Models were developed for the following cancer groups, plus all cancers: leukaemia/lymphoma (12 predictors), central nervous system tumours (seven predictors), bone tumours and soft tissue sarcomas (four predictors), abdominal tumours (seven predictors) Predictors, adjusted ORs (95% CIs) and PPVs (95% CIs), respectively, for all cancers' model: 1. Pallor: 83.7 (18.0, 390.5); 0.41% (0.12%, 1.34%) 2. Head and neck lump mass swelling: 16.9 (5.2, 54.9); 0.30% (010%, 0.84%) 3. Lump mass swelling: 21.8 (9.6, 49.7); 0.11% (0.06%, 0.20%) 4. Lymphadenopathy: 10.1 (5.9, 17.4); 0.09% (0.06%, 0.13%) 5. Abnormal movement: 16.4 (7.8, 34.9); 0.08% (0.04%, 0.14%) 6. Bruising: 12.3 (5.5, 27.8); 0.08% (0.05%, 0.13%)	Twelve features of childhood cancers were identified, each of which increased the risk of cancer at least 10-fold

continued

TABLE 40 Systematic review 2: included studies, results (*continued*)

Study	Model evaluation	Results	Interpretation and discussion
		7. Fatigue: 7.7 (3.8, 15.8); 0.07% (0.04%, 0.12%) 8. Bleeding: 9.9 (4.9, 20.2); 0.06% (0.03%, 0.10%) 9. Headache: 6.1 (3.8, 9.9); 0.06% (0.04%, 0.08%) 10. Visual: 10.4 (4.4, 24.3); 0.06% (0.03%, 0.10%) 11. Pain: 7.3 (4.0, 13.4); 0.04% (0.03%, 0.06%) 12. Musculoskeletal symptoms: 5.3 (3.6, 7.7); 0.04% (0.03%, 0.07%) When each of these 12 symptoms was combined singly with at least three consultations in a 3-month period, the probability of cancer was between 11 and 76 in 10,000	
Muris 1995[56]	Apparent performance	Statistically significant predictors [adjusted ORs (95% CI)] were: • No specific character to pain: 5.70 (1.97 to 16.51) • Weight loss: 4.36 (1.72 to 11.11) • ESR of > 20 mm/hour: 3.00 (1.10 to 8.17) • Male sex: 2.37 (1.20 to 6.99) • Greater age (years): 1.08 (1.04 to 1.11)	The aim of the study was to identify predictors for organic disease, with a secondary analysis looking at predictors for neoplasms
Elias 2017[50]	External validation	Elias *et al.*[50] ranked the Muris *et al.*[56] model as 13th out of 19 models evaluated	The top-ranked model was the NICE guidelines. Note that Elias 2017 used a very broad definition of prediction model, which included guidelines and weighted scores

AUC, area under the curve; BMI, body mass index; BTC, biliary tract cancer; CART, classification and regression trees; CEDAR, Cost-Effectiveness of a Decision rule for Abdominal complaints in Primary care; ESR, erythrocyte sedimentation rate; GI, gastrointestinal; HR, hazard ratio; IBD, inflammatory bowel disease; LR, likelihood ratio; MCV, mean corpuscular volume; OR, odds ratio; PDAC, pancreatic ductal adenocarcinoma; RF, random forest; SVM, support vector machine; WBC, white blood cell.

TABLE 41 Systematic review 2: risk-of-bias assessment, questions 1–3

Study	1a. Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	1b. Were all inclusions and exclusions of participants appropriate?	1c. Were participants enrolled at a similar state of health, or were predictors considered to account for differences?	I. Participant selection	2a. Were predictors defined and assessed in a similar way for all participants in the study?	2b. Were predictor assessments made without knowledge of outcome data?	2c. Are all predictors available at the time the model is intended to be used?	2d. Were all relevant predictors analysed?	II. Predictors	3a. Was a prespecified outcome definition used?	3b. Were predictors excluded from the outcome definition?	3c. Was the outcome defined and determined in a similar way for all participants?	3d. Was the outcome determined without knowledge of predictor information?	III. Outcome
Marshall 2011[22]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Unclear	Unclear risk	Yes	Yes	Yes	Yes	Low risk
McCowan 2011[60]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Dommett 2013[64]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Keane 2014[62]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Fijten 1995[55]	No	Yes	Yes	High risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Hodder 2005[77]	No	Unclear	Yes	High risk	Yes	Yes	Yes	Unclear	Unclear risk	Yes	No	Yes	No	High risk
Kop 2015[61]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Unclear	Unclear risk	Yes	Yes	Yes	Yes	Low risk
Hippisley-Cox 2013[53]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Hippisley-Cox 2013[54]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Collins 2012[15]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Hippisley-Cox 2012[80]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Collins 2013[88]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Hippisley-Cox 2011[82]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Hippisley-Cox 2011c[52]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Collins 2013[89]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Hippisley-Cox 2011[84]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Collins 2013[14]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Hippisley-Cox 2012[32]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Collins 2013[91]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk

continued

TABLE 41 Systematic review 2: risk-of-bias assessment, questions 1–3 (*continued*)

Study	1a. Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	1b. Were all inclusions and exclusions of participants appropriate?	1c. Were participants enrolled at a similar state of health, or were predictors considered to account for differences?	I. Participant selection	2a. Were predictors defined and assessed in a similar way for all participants in the study?	2b. Were predictor assessments made without knowledge of outcome data?	2c. Are all predictors available at the time the model is intended to be used?	2d. Were all relevant predictors analysed?	II. Predictors	3a. Was a prespecified outcome definition used?	3b. Were predictors excluded from the outcome definition?	3c. Was the outcome defined and determined in a similar way for all participants?	3d. Was the outcome determined without knowledge of predictor information?	III. Outcome
Hippisley-Cox 2012[85]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Hamilton 2015[86]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Shephard 2015[68]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Shephard 2012[69]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Stapley 2017[72]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Hamilton 2007[87]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Hamilton 2005[81]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Hamilton 2009[74]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Stapley 2013[71]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Hamilton 2005b[51]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Hamilton 2009b[83]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Stapley 2012[70]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Hamilton 2006[90]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Walker 2013[73]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Shephard 2016[67]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Shephard 2013[66]	Yes	Yes	Yes	Low risk	Yes	Unclear	Yes	Yes	Unclear risk	Yes	Yes	Yes	Yes	Low risk
Iyen-Omofoman 2013[21]	Yes	Yes	Yes	Low risk	Yes	Unclear	Yes	Yes	Unclear risk	Yes	Yes	Yes	Yes	Low risk
Elias 2017[50]	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk
Nørrelund 1996[57]	Yes	Yes	Yes	Low risk	Yes	Unclear	Yes	Unclear	Unclear risk	Yes	Yes	Yes	Unclear	Low risk
Muris 1995[56]	Yes	Unclear	Yes	Unclear risk	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Low risk

TABLE 42 Systematic review 2: risk-of-bias assessment, questions 4 and 5

Study	4a. Were there a reasonable number of outcome events?	4b. Was the time interval between predictor assessment and outcome determination appropriate?	4c. Were all enrolled participants included in the analysis?	4d. Were participants who were missing data handled appropriately?	IV. Sample size and participant flow	5a. Were non-binary predictors handled appropriately?	5b. Was selection of predictors based on univariable analysis avoided?	5c. Was model overfitting (optimism in model performance) accounted for, for example using bootstrapping or shrinkage techniques?	5d. Were any complexities in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?	5e. Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	5f. For the model or any simplified score, were relevant performance measures evaluated, for example calibration, discrimination, (re)classification and net benefit?	5g. Was the model recalibrated or was it likely (based on the evidence presented, e.g. calibration plot) that recalibration was not needed?	V. Analysis
Marshall 2011[22]	Yes	Yes	Yes	Unclear	Unclear risk	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Low risk
McCowan 2011[60]	No	Yes	No	Yes	High risk	Yes	No	No	Yes	Yes	Yes	No	High risk
Dommett 2013[64]	Yes	Yes	Yes	Yes	Low risk	Yes	No	Unclear	Yes	Yes	Yes	Unclear	Unclear risk
Keane 2014[62]	Yes	Yes	Yes	Unclear	Unclear risk	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Low risk
Fijten 1995[55]	Unclear	Yes	Yes	Yes	Unclear risk	Yes	No	No	Yes	Yes	Yes	No	High risk
Hodder 2005[77]	Yes	Yes	Yes	Yes	Low risk	Unclear	No	No	Yes	Yes	Yes	No	Unclear risk
Kop 2015[61]	Yes	Yes	Yes	Unclear	Unclear risk	Yes	Yes	Unclear	Unclear	Unclear	Unclear	Unclear	Unclear risk
Hippisley-Cox 2013[53]	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Low risk
Hippisley-Cox 2013[54]	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Low risk
Collins 2012[15]	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Low risk
Hippisley-Cox 2012[80]	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Low risk
Collins 2013[88]	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Low risk
Hippisley-Cox 2011[82]	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Low risk
Hippisley-Cox 2011[52]	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Low risk
Collins 2013[89]	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Low risk
Hippisley-Cox 2011[84]	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Low risk
Collins 2013[14]	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Low risk
Hippisley-Cox 2012[32]	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Low risk
Collins 2013[91]	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Low risk

continued

TABLE 42 Systematic review 2: risk-of-bias assessment, questions 4 and 5 (*continued*)

Study	4a. Were there a reasonable number of outcome events?	4b. Was the time interval between predictor assessment and outcome determination appropriate?	4c. Were all enrolled participants included in the analysis?	4d. Were participants who were missing data handled appropriately?	IV. Sample size and participant flow	5a. Were non-binary predictors handled appropriately?	5b. Was selection of predictors based on univariable analysis avoided?	5c. Was model overfitting (optimism in model performance) accounted for, for example using bootstrapping or shrinkage techniques?	5d. Were any complexities in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?	5e. Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	5f. For the model or any simplified score, were relevant performance measures evaluated, for example calibration, discrimination, (re)classification and net benefit?	5g. Was the model recalibrated or was it likely (based on the evidence presented, e.g. calibration plot) that recalibration was not needed?	V. Analysis
Hippisley-Cox 2012[85]	Yes	Yes	Yes	Yes	Low risk	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Low risk
Hamilton 2015[86]	Yes	Yes	Yes	Unclear	Unclear risk	Yes	No	Unclear	Yes	Yes	No	Unclear	High risk
Shephard 2015[68]	Yes	Yes	Yes	Unclear	Unclear risk	Yes	No	No	Yes	Yes	No	No	High risk
Shephard 2012[69]	Yes	Yes	Yes	Unclear	Unclear risk	Yes	No	No	Yes	Yes	No	No	High risk
Stapley 2017[72]	Yes	Yes	Yes	Unclear	Unclear risk	Yes	No	No	Yes	Yes	No	No	High risk
Hamilton 2007[87]	Yes	Yes	Yes	Unclear	Unclear risk	Yes	No	Unclear	Yes	Yes	No	Unclear	High risk
Hamilton 2005[81]	Yes	Yes	Yes	Unclear	Unclear risk	Yes	No	Unclear	Yes	Yes	No	Unclear	High risk
Hamilton 2009[74]	Yes	Yes	Yes	Unclear	Unclear risk	Yes	No	Unclear	Yes	Yes	No	Unclear	High risk
Stapley 2013[71]	Yes	Yes	Yes	Unclear	Unclear risk	Yes	No	No	Yes	Yes	No	No	High risk
Hamilton 2005[51]	Yes	Yes	Yes	Unclear	Unclear risk	Yes	No	Unclear	Yes	Yes	No	Unclear	High risk
Hamilton 2009[83]	Yes	Yes	Yes	Unclear	Unclear risk	Yes	No	Unclear	Yes	Yes	No	Unclear	High risk
Stapley 2012[70]	Yes	Yes	Yes	Unclear	Unclear risk	Yes	No	No	Yes	Yes	No	No	High risk
Hamilton 2006[90]	Yes	Yes	Yes	Unclear	Unclear risk	Yes	No	Unclear	Yes	Yes	No	Unclear	High risk
Walker 2013[73]	Yes	Yes	Yes	Unclear	Unclear risk	Yes	No	No	Yes	Yes	No	No	High risk
Shephard 2016[67]	Yes	Yes	Yes	Unclear	Unclear risk	Yes	No	No	Yes	Yes	No	No	High risk
Shephard 2013[66]	Yes	Yes	Yes	Unclear	Unclear risk	Unclear	No	No	Unclear	Yes	No	No	High risk
Iyen-Omofoman 2013[21]	Yes	Yes	Yes	Unclear	Unclear risk	Unclear	No	No	Unclear	Yes	No	No	High risk
Elias 2017[50]	Yes	Unclear	No	Unclear	Unclear risk	Unclear	N/A	N/A	N/A	N/A	Yes	Unclear	Unclear risk
Nørrelund 1996[57]	Yes	Yes	Unclear	Unclear	Unclear risk	Unclear	Yes	No	No	No	Yes	No	High risk
Muris 1995[56]	No	Yes	Unclear	Unclear	Unclear risk	Yes	No	No	Unclear	Yes	No	No	High risk

N/A, not applicable.

TABLE 43 Studies excluded from the review but highlighted for interest: studies with secondary care (i.e. referred or diagnosed) populations

Study	Prediction tool	Cancer type(s)	Country	Population
Adelstein 2010[204]	No name (multiple variable model for CRC)	Colorectal	Australia	Patients referred for colonoscopy, NSW Australia
Baicus 2006[205]	No name (generic cancer)	Multiple	Romania	Patients admitted to a hospital in Romania (January–September 2003) with involuntary weight loss
Ewing 2016[206]	No name (Swedish – non-metastatic CRC)	Colorectal	Sweden	Patients with cancer
Galvin 2014[207]	Prediction rule for breast cancer	Breast	Ireland	Prospective cohort of consecutive patients with breast cancer symptoms reviewed at the symptomatic breast units in Beaumont hospital (Ireland)
Khademi 2012[208]	No name (risk-prediction model for upper GI cancer)	GI	Islamic Republic of Iran	Patients diagnosed with dyspepsia and referred to a tertiary referral gastroenterology clinic in Tehran (Islamic Republic of Iran) from 2002 to 2009
Moore 2011[209]	Risk of Ovarian Malignancy Algorithm	Ovarian	USA	Premenopausal and postmenopausal women aged ≥ 18 years presenting to a generalist (defined as a general gynaecologist, internist, family practitioner gastroenterologist or general surgeon) with an ovarian cyst or an adnexal mass and subsequently scheduled to undergo surgery
Saraiva 2016[210]	No name (uses of AI to aid diagnosis of GI cancers)	GI	Brazil	Patients with GI cancer
Reeves 2003[211]	BREASTAID	Breast	USA	Referred and follow-up patients
Robertson 2006[212]	No name	Colorectal	UK	Patients referred with rectal bleeding
Simpkins 2017[213]	No name	Colorectal	USA	Patients newly referred with GI symptoms from primary care to two secondary care centres
Koning 2015[214]	No name	Colorectal	The Netherlands	Patients referred for colonoscopy
Kim 2009[215]	No name	Ovarian	The Republic of Korea	All referred/screened patients
Høgdall 2011[216]	No name	Ovarian	Denmark	Complex laboratory equipment required – Biomek® 2000 (Beckman Coulter Inc., Brea CA, USA); primary care availability unlikely

AI, artificial intelligence; GI, gastrointestinal.

TABLE 44 Studies excluded from the review but highlighted for interest: studies not on prediction models

Study	Prediction tool	Cancer type(s)	Country	Comment
Bourne 2012[23]	BLINCK	Melanoma	Australia	Algorithm (clinical and dermatoscopic criteria)
Dolianitis 2005[217]	Dermoscopic algorithms: seven-point checklist, the ABCD rule and the Menzies method[218]	Melanoma	Australia	Algorithms
Emery 2010[219]	PCSA	Melanoma	UK/ Australia	Does not calculate risk of cancer (score)
Gerbert 2000[220]	Decision support software	Melanoma	CA, USA	Algorithm (clinical and dermatoscopic criteria)
Rogers 2016[221]	TADA	Melanoma	NY, USA	Algorithm (dermatoscopic criteria)
Walter 2012[222]	MoleMate™ (MedX Health Corp., Mississauga, ON, Canada), seven-point checklist	Melanoma	UK	MoleMate is a 'computerised diagnostic tool' involving a novel imaging technique (SIA); seven-point checklist is an algorithm
Walter 2013[223]	Original and weighted seven-point checklist	Melanoma	UK	Does not calculate risk of cancer (score) diagnosis of pigmented skin lesions (including melanoma) in primary care; based on RCT reported in Walter 2012,[222] record 6175, currently included in SR1
Hodder 2005[77]	Netherlands model; Harvard model; weighted numerical score	CRC	UK	Develop a system to compare and validate referral guidelines (is this SR2 external validation rather than SR1-impact assessment?)
Ballal 2010[224]	Weighted numerical score	CRC	Wales (UK)	Does not calculate risk of cancer (score) external validation of score developed by Selvachandran 2002[225]
Selvachandran 2002[225]	Weighted numerical score	CRC	UK	Does not calculate risk of cancer (score)
Smith 2006[226]	Weighted numerical score	CRC	UK	Does not calculate risk of cancer (score)
Zortea 2014[227]	No name (computer image processing algorithm)	Melanoma	Germany, Norway	Image processing system for skin cancer detection
Grewal 2013[228]	No name (scoring system derived from Hamilton 2009 *et al.*[83] data)	Ovarian	UK	Scoring system derived from Hamilton 2009[83] data
Shahzad 2015[229]	No name	Ovarian	Pakistan	Index score
Rossing 2010[230]	No name	Ovarian	USA	Index score
Lim 2012[231]	No name	Ovarian	UK	Scoring system
Law 2014[232]	No name	Colorectal	Malaysia	Index score derived from Selvachandran 2002[225] data

ABCD, asymmetry, borders, colours, differential structure components; BLINCK, Benign, Lonely, Irregular, Nervous, Change, Known clues; PCSA, primary care scoring algorithm; SIA, spectrophotometric intracutaneous analysis; TADA, triage amalgamated dermoscopic algorithm.

TABLE 45 Systematic review 2 (updated searches): included studies, study characteristics (1)

Study	Prediction model	Cancer type(s)	Country	Setting	Study design	Stage of development	Data source
Holtedahl 2018[63]	Prediction model for abdominal cancers	Abdominal [including all cancers of digestive organs, female genital organs, urinary organs (including testis)]. Other cancers were included in additional analyses if they were associated with abdominal signs or symptoms	Norway, Denmark, Sweden, Scotland, Belgium and the Netherlands	Primary care	Prospective cohort	Apparent performance	GP records

TABLE 46 Systematic review 2 (updated searches): included studies, study characteristics (2)

Study	Participants	Candidate predictors	Outcome to be predicted	Sample size	Number/handling of missing data
Holtedahl 2018[63]	493 GPs recruited via Cancer and Primary Care Research International Network, February–July 2011 GPs recorded consultations over a period of 10 days for patients aged ≥ 16 years. If abdominal symptoms were mentioned in the consultation, specific symptom-related questions were asked using a pro forma GPs also identified all those with a diagnosis of abdominal cancer within 6 months after the GP survey, regardless of whether or not the patient attended during the survey period	Abdominal pain (lower and upper), constipation, diarrhoea, distended abdomen/bloating, increased belching/flatulence, acid regurgitation, rectal bleeding, unexpected genital bleeding, macroscopic haematuria, increased urinary frequency, other abdominal problems	Cancer diagnosis within 180 of GP survey. Diagnosis taken from GP records 8 months after GP survey. Study authors distinguished between abdominal and non-abdominal cancers	61,802 patients, including 175 cases (0.28%)	Not reported

TABLE 47 Systematic review 2 (updated searches): included studies, model development and performance

Study	Model development	Model performance
Holtedahl 2018[63]	Cox proportional hazards models. Report univariate and multivariate analyses (for the most frequent symptoms and combinations of symptoms adjusted for sex) Authors report no evidence to reject proportional hazards assumption Main analyses included all patients with new abdominal cancer diagnosis within 180 days. Other analyses looked at all cancers, or abdominal cancers also beyond 180 days 0.05 level of statistical significance used	Sensitivity, specificity, likelihood ratios and PPVs reported for individual symptoms and combinations of symptoms (by age group or sex)

TABLE 48 Systematic review 2 (updated searches): included studies, results

Study	Model evaluation	Results	Interpretation and discussion
Holtedahl 2018[63]	Developmental data set only	HR (95% CI): • Abdominal pain (upper, single symptom): 4.8 (1.9 to 11.8) • Abdominal pain (lower, single symptom): 5.8 (2.4 to 14.3) • Constipation (single symptom): 6.8 (2.1 to 21.8) • Rectal bleeding (single symptom): 19.1 (8.7 to 41.7) • Any other single symptoms (grouped): 4.7 (2.8 to 7.9) • Two abdominal symptoms 4.6 (2.5 to 8.5) • Three or more abdominal symptoms: 14.0 (9.1 to 21.6) HRs (95% CIs) were also reported for combinations of symptoms – see table 4 of Holtedahl 2018[63]	The strength of the association between abdominal symptoms and cancers highlights the importance of responding to these symptoms. But, as some new cancers did not involve these symptoms, clinical suspicion is needed

TABLE 49 Systematic review 2 (updated searches): risk-of-bias assessment, questions 1–3

Study	1a. Were appropriate data sources used, e.g. cohort, RCT or nested case-control study data?	1b. Were all inclusions and exclusions of participants appropriate?	1c. Were participants enrolled at a similar state of health, or were predictors considered to account for differences?	I. Participant selection	2a. Were predictors defined and assessed in a similar way for all participants in the study?	2b. Were predictor assessments made without knowledge of outcome data?	2c. Are all predictors available at the time the model is intended to be used?	2d. Were all relevant predictors analysed?	II. Predictors	3a. Was a prespecified outcome definition used?	3b. Were predictors excluded from the outcome definition?	3c. Was the outcome defined and determined in a similar way for all participants?	3d. Was the outcome determined without knowledge of predictor information?	III. Outcome
Holtedahl 2018[63]	Yes	Yes	Unclear	Unclear risk	Yes	Yes	Yes	Yes	Low risk	Unclear	Yes	Unclear	Yes	Unclear risk

TABLE 50 Systematic review 2 (updated searches): risk-of-bias assessment, questions 4 and 5

Study	4a. Were there a reasonable number of outcome events?	4b. Was the time interval between predictor assessment and outcome determination appropriate?	4c. Were all enrolled participants included in the analysis?	4d. Were participants who were missing data handled appropriately?	IV. Sample size and participant flow	5a. Were non-binary predictors handled appropriately?	5b. Was selection of predictors based on univariable analysis avoided?	5c. Was model overfitting (optimism in model performance) accounted for, for example using bootstrapping or shrinkage techniques?	5d. Were any complexities in the data (e.g. competing risks, multiple events per individual) accounted for appropriately?	5e. Do predictors and their assigned weights in the final model correspond to the results from multivariable analysis?	5f. For the model or any simplified score, were relevant performance measures evaluated, for example calibration, discrimination, (re)classification and net benefit?	5g. Was the model recalibrated or was it likely (based on the evidence presented, e.g. calibration plot) that recalibration was not needed?	V. Analysis
Holtedahl 2018[63]	Yes	Unclear	No	Unclear	High risk	Unclear	Yes	Unclear	Yes	Yes	Yes	No	Unclear risk

Appendix 4 Updated review: supplementary tables

TABLE 51 Systematic review 2: updated searches[a]

Author and year; country	Study aim	Patient population (n analysed)	Study design; [data sources]	Interval(s) definition (T interpretation), and how interval was analysed	Outcome measures	Comments and method of analysis	Addressed waiting time paradox?
Breast							
Ángeles-Llerenas 2016;[100] Mexico	To estimate the effect of care delivery delays on survival among women with breast cancer	Patients with histopathologically diagnosed breast cancer (2007–9) attending hospital for treatment ($n = 854$)	• Retrospective cohort (multicentre) • [Patient interviews]	• Symptom onset to treatment (T5); • Symptom onset to first consultation with a doctor about symptoms (T1) • Receipt of mammography results to diagnostic biopsy results • Biopsy to the initiation of treatment (T15) • Days; time divided into quartiles with Q1 as reference	Survival (from initiating treatment; 5 years' follow-up)	Cox proportional risk model adjusting for, among others, stage and sign and symptoms. Analysis also stratified by stage	No evidence of this
Murchie 2015;[102] Scotland, UK	To explore whether or not longer provider delays (between first presentation and treatment) were associated with later stage and poorer survival in women with symptomatic breast cancer	Symptomatic patients diagnosed with breast cancer (1997–8) in Northern Scotland identified via linked data sets ($n = 850$)	• Retrospective cohort (multicentre). • [Linked data sets: primary care data, Cancer Registry, Death Registry and the acute hospital discharge]	First presentation to primary care to first treatment or to definitive diagnosis (T9) Weeks; continuous variable, and modelled as four periods: 4 (reference), 13, 26, and 39 corresponding to knots placed on distribution of interval	• Survival (all-cause survival from presentation to primary care) • Stage (III or IV)	• Analysis of survival based on 850 patients, and stage based on 777 patients • Restricted cubic spline curves. OR (99% CI) obtained from logistic regression (stage: III/IV vs. I/II) or Cox survival models using sequential adjustment for patient and tumour factors (including, among others, grade, and signs and symptoms)	No evidence of this

Author and year; country	Study aim	Patient population (n analysed)	Study design; [data sources]	Interval(s) definition (T interpretation), and how interval was analysed	Outcome measures	Comments and method of analysis	Addressed waiting time paradox?
Pace 2015;[103] Rwanda	To investigate the magnitude, impact of and risk factors for, patient and system delays in breast cancer diagnosis in Rwanda	Women who presented with a breast complaint to oncology clinic (2012–14) and abnormal breast examination findings ($n = 144$)	• Retrospective cohort (two rural sites) • [Interview-questionnaire and medical records]	• Symptom development to patient's first presentation to a health-care provider (T1) • First presentation to diagnosis (T8) • Months; divided into four groups: < 3 (reference), ≥ 3 to < 6, ≥ 6 to 12, and > 12 months	Stage (advanced)	Multivariate ordinal logistic regression adjusting for clinical and demographic characteristics (did not incorporate symptoms or tumour grade) used to obtain OR (95% CI) (presented as a forest plot only)	No evidence of this
Unger-Saldaña, 2015;[104] Mexico	To determine the correlation between health system delay and clinical disease stage in patients with breast cancer	• Patients referred to one of the four largest public cancer hospitals in Mexico City for the evaluation of probable breast cancer (patients interviewed 2009–11) • (Included symptomatic patients and those identified via screening) • [$n = 886$; 670 (76%) symptomatic]	• Cross-sectional study (multicentre) • [Patient interview questionnaires and patients' hospital records]	'Patient interval': identification of the problem (via symptoms or screening) to first medical consultation (T3) 'Health system interval': first medical consultation to first treatment (T14) Months; continuous variable	Stage (advanced)	• Cox regression model used to evaluate the significance of differences in interval lengths between different stages (stage 0–I, II, III, IV, unknown); analysis did not control for whether the problem had been identified via screening or symptoms. Logistic regression model then used to estimate the average marginal effects of 'patient' and 'health system' interval on probability of being diagnosed with advanced breast cancer (stages III and IV). This analysis was adjusted for the means of problem identification (symptoms vs. screening) as well as five other patient characteristics) • Outliers (total delay ≥ 50 months) were excluded. Data on stage only available for 597 patients who received treatment at participating hospitals	No evidence of this. Patients with delay of ≥ 50 months excluded

continued

TABLE 51 Systematic review 2: updated searches[a] (*continued*)

Author and year; country	Study aim	Patient population (n analysed)	Study design; [data sources]	Interval(s) definition (T interpretation), and how interval was analysed	Outcome measures	Comments and method of analysis	Addressed waiting time paradox?
Bladder							
Bryan 2015;[105] England, UK	To investigate the influence of tumour factors, patient factors, carcinogen exposure and pathway delays on the long-term outcome of urothelial bladder cancer	Patients newly diagnosed with urothelial bladder cancer (1991–2) in the West Midlands region ($n = 1478$)	• Retrospective cohort (multicentre) • [Data collected at 'recruitment', and cancer registry notified (cohort followed up prospectively for survival)]	• Symptoms onset to first treatment (surgery) (T5) • First GP referral to first treatment (T12) • Symptoms onset to first GP referral to secondary care (T2) • GP referral to first hospital attendance for urological assessment (T10) • First hospital attendance to first treatment (T14) • Days; divided into long and short intervals for each T: T5 $\leq$ 110 vs. > 110; T12 $\leq$ 68 vs. > 68; T2 $\leq$ 14 vs. > 14; T10 $\leq$ 28 vs. > 28; T14 $\leq$ 20 vs. > 20 days	• Survival (from treatment; 17 years' follow-up) • Stage (advanced) • Tumour size (> 2 cm)	Cox regression used to identify factors associated with longer intervals, including, among others, stage and tumour size. Cox regression also used to assess predictors of poor survival, including delays, tumour stage, grade, and size	No evidence of this
Colorectal							
Aslam 2017;[110] England, UK	To determine whether or not referral and diagnosis through the 2WW pathway confers a survival advantage on these patients	Patients diagnosed with CRC at local university hospitals (2005–12) and undergoing surgery (study included patients presenting via screening, but not in the analysis of OS; analysis also conducted excluding emergency presentations) ($n = 2604$)	• Retrospective comparative cohort (single NHS trust) • [Hospital database on CRC outcomes]	Referral to initial treatment (T12) Days; Kaplan–Meier curves developed for two groups: < 62 vs. > 62 days	Survival (median OS)	Kaplan–Meier methods (log-rank test) • Compared referral type groups (2WW, urgent, routine, emergency and screening pathways) and referral interval (within and after 62 days, excluding emergency presentations) • Analysis of T12 for 2WW, urgent, and routine referrals only based on 2335 patients	Excluded emergency presentations

Author and year; country	Study aim	Patient population (n analysed)	Study design; [data sources]	Interval(s) definition (T interpretation), and how interval was analysed	Outcome measures	Comments and method of analysis	Addressed waiting time paradox?
Chen 2017;[112] USA	To examine patient presentation, provider evaluation and time to diagnosis, which can affect stage and prognosis. Also compared the symptoms, clinical features, time to diagnosis and stage of young-onset (aged < 50 years) CRC with those of patients diagnosed at age ≥ 50 years	Patients seen at a cancer institute (2008–14), with pathologically confirmed CRC diagnosis; included two separate cohorts: those aged < 50 years and those aged ≥ 50 years (included emergency presentations) ($n = 458$)	• Retrospective cohort study (single centre) • [Cancer institute's research database (clinical records)]	Symptom onset to diagnosis (defined as the sum of 'symptom duration' and 'workup duration') (T4) Symptom duration: symptom onset to initial visit (T1) Workup duration: first medical visit to date of pathological diagnosis (T8) Days; median intervals compared for advanced and non-advanced stage at diagnosis (but not assessed statistically)	Stage (III and IV)	• 253 had young-onset CRC, and 232 were diagnosed at age ≥ 50 years. First medical visit was primary care for 320 (70%) patients, and emergency department for 113 (25%) patients • Stage stratified as advanced (III or IV) and non-advanced (Stage I or II). No statistical analysis conducted. When comparing patients with advanced stage cancer with those with non-advanced stage cancer, the difference in median T1 in young-onset CRC was 30 days, and for T4 it was 40 days. It was ≤ 14 days for all other interval comparisons • Assessed extent of intervals, compared intervals in younger and older cohorts, assessed factors associated with intervals, and whether or not longer periods were associated with more extensive disease	Did look at delay to diagnosis by CRC symptoms vs. non-specific symptoms, but do not account for this in the main analysis of delay vs. stage at diagnosis

continued

TABLE 51 Systematic review 2: updated searches[a] (*continued*)

Author and year; country	Study aim	Patient population (n analysed)	Study design; [data sources]	Interval(s) definition (T interpretation), and how interval was analysed	Outcome measures	Comments and method of analysis	Addressed waiting time paradox?
Helewa 2013;[106] Canada	To determine if increased wait times for CRC treatments resulted in worse survival outcomes. Secondary objectives: to determine predictors of longer waiting times and urgent presentations	Patients diagnosed with stage I–IV CRC, (2004–6) who underwent major surgical resections. Included emergency presentations, which were also analysed separately ($n = 1628$)	• Retrospective population-based cohort study (multicentre) • [Cancer Registry linked with other administrative databases]	• Total wait time: patient's index contact to first treatment (T9) (defined as the sum of the 'diagnostic wait time' and 'treatment wait time') • Diagnostic wait time: index contact to date of diagnosis (T8) • Treatment wait time: date of diagnosis to first treatment (T15) • Days; continuous variable, and divided into quartiles: 0 to ≤ 43 (reference), > 43 to ≤ 95, > 95 to ≤ 166, > 166 to ≤ 513 days; Kaplan–Meier curves for each quartile developed	Survival (5-year OS)	• Only 1307 out of 1628 patients contributed to the analysis of T9 • HR (95% CI) obtained from multivariate Cox proportional hazards model using bootstrap resampling, adjusted for stage, emergency presentations, location and other factors	Adjusted for emergency presentations
Janssen 2016;[114] Canada	To evaluate local wait times for colonoscopy in patients who were subsequently found to have CRC	Patients referred to and subsequently seen by the gastroenterologist and had a pathological diagnosis of CRC (2010–13). (Excluded emergency presentations and those seen as inpatients.) ($n = 246$)	• Retrospective cohort study (single centre) • [Hospital pathology database, patient records]	Primary care referral to definitive endoscopy (T11) Days; divided into 2 groups: < 60 vs. > 60 days	Node positivity, and presence of distant metastases	Assessed whether or not local wait times (from referral) were in accordance with guidelines, but also evaluated the effect of wait time on the presence of node positivity, and distant metastases at diagnosis. Compared the outcomes of patients who were seen within the guidelines with the outcomes of patients seen outside the guidelines using Fisher's exact or chi-squared tests	Excluded emergency presentations and inpatients. Majority of patients had alarm symptoms at referral

Author and year; country	Study aim	Patient population (n analysed)	Study design; [data sources]	Interval(s) definition (T interpretation), and how interval was analysed	Outcome measures	Comments and method of analysis	Addressed waiting time paradox?
Leiva 2017;[113] Spain	To determine the diagnostic intervals for the diagnosis of CRC by using three different sources of information and to analyse the relationship of CRC stage with the diagnostic interval according to the source of information	Patients diagnosed with CRC at one of nine public hospitals (2006–8) and registered with GP. (Included emergency presentation.) ($n = 715$)	• Retrospective population-based study (multicentre) • [Patient interviews, hospital records, GP records]	Diagnostic interval calculated using three different information sources: patient recall, hospital records and GP records For hospital and GP records, diagnostic interval defined as first registry of symptoms to diagnosis (T8) For patient-recall, diagnostic interval defined as onset of symptoms to diagnosis (T4) Days; median interval estimated for each TNM stage (classified as 0, 1, 2, 3, 4, unknown), and cumulative distribution of the three intervals also plotted for four stage groupings: 0–I, II–III, IV, unknown	Stage (more advanced)	• Time to diagnosis and stage of symptomatic CRC determined by three different sources of information. Analysis of 'patient recall' included 715 patients, 'hospital records' 637, and 'GP records' 316 • Assessed association between diagnostic intervals (stratified by information source) and various patient and clinical characterises (including, among others, stage, emergency presentation, location, symptoms at presentation, and perceived seriousness of symptoms; univariate analyses). Kaplan–Meier methods used to assess independent predictors of shorter interval using Cox regression and then multivariate Cox proportional hazards model with time-dependent covariates	No evidence of this. Main analyses of interest were unadjusted, and emergency presentations were included

continued

TABLE 51 Systematic review 2: updated searches[a] (*continued*)

Author and year; country	Study aim	Patient population (n analysed)	Study design; [data sources]	Interval(s) definition (T interpretation), and how interval was analysed	Outcome measures	Comments and method of analysis	Addressed waiting time paradox?
Murchie 2014;[107] Scotland, UK	To explore if longer provider delays (time from presentation to treatment) were associated with more advanced stage disease at diagnosis and poorer survival	Symptomatic patients diagnosed with CRC (1997–8) in northern Scotland identified via linked data sets (data collected in 2000–1). (Included emergency presentation.) ($n = 958$)	• Retrospective population-based cohort study (multicentre) • [Linked data sets: primary care data, Cancer Registry, Death Registry, and the acute hospital discharge]	'Provider delay: first presentation in primary care to treatment or definitive diagnosis (T9) Weeks; continuous variable, and modelled as four periods: 4 (reference), 20, 40 and 60 corresponding to knots placed on distribution of interval	• Stage (advanced; C and D) • Survival (from date of presentation; 2 years' follow-up)	• Analysis of survival based on 958 patients, and of stage based on 868 patients • Restricted cubic spline curves. OR (99% CI) obtained from logistic regression (stage) or Cox survival models using sequential adjustment for patient and tumour factors (including, among others, grade, emergency admission, and signs and symptoms)	Models adjusted for emergency admission, signs and symptoms
Patel, 2018;[111] Scotland, UK	To determine the impact of compliance with the 62-day pathway on outcomes in patients with CRC in the long term	Patients treated for CRC at one of three district general hospitals (1999–2005). Included patients presenting electively or in an emergency setting (with patients who did not go on to have surgery within 72 hours of admission reclassified as elective). ($n = 1012$; 794 elective, 218 emergency)	• Retrospective cohort (multicentre) • [Patient records]	Initial GP referral to first treatment (T12) Days: dichotomised based on being seen within 62-day standard or not; divided into four groups: standard met (elective), standard met (emergency), standard failed (elective) and standard failed (emergency); also plotted separate Kaplan–Meier curves for ≤ 62 vs. > 62 days	Survival (mean OS from diagnosis based on ≥ 6 years' follow-up); and stage (C or D vs. A or B)	• Statistical analysis based on log-rank and chi-squared tests. Comparison of Kaplan–Meier curves for patients seen within 62 days or not • Stratified analysis according to whether patients received emergency or elective surgery. Also analysed 30-day postoperative mortality and cause of death	Excluded emergency presentations for survival outcomes
Pita-Fernández 2016;[109] Spain	To analyse the relationship between the interval from first symptom to diagnosis and survival in CRC	Patients diagnosed with CRC at a university hospital (1994–2000) ($n = 942$)	• Retrospective cohort study (single centre) • [Hospital records]	'Diagnostic delay': onset of symptoms to diagnosis (biopsy or direct surgery) (T4) Months; analysed as a continuous variable (using penalised splines), and quartiles: < 1.5, 1.5 to 3.4, 3.4 to 6.4, > 6.4 months, with Kaplan–Meier curves developed and compared for each quartile	Survival (5 years, from diagnosis)	Kaplan–Meier curves (log-rank test). Cox proportional hazards model with restricted cubic spline curves. Sequential adjustment for stage (initial model adjusted for age and sex). Stratified analysis colon and rectal	No evidence of this

Author and year; country	Study aim	Patient population (n analysed)	Study design; [data sources]	Interval(s) definition (T interpretation), and how interval was analysed	Outcome measures	Comments and method of analysis	Addressed waiting time paradox?
Lung							
Gildea 2017;[118] USA	To assess the wait time to diagnose NSCLC and the cost of diagnosis and treatment based on the stage at diagnosis	Patients diagnosed with NSCLC (2007–11) with both medical and pharmacy benefits (health insurance) ($n = 1210$)	• Retrospective cohort study (multicentre) • [Cancer registry linked to health insurance claims database]	First diagnostic test for lung cancer to a definitive diagnosis (T13) Months; comparison of mean (SD) interval for each stage (I, II, IIIa, IIIb, IV); 'long mean delay': 5 or 6 months	Stage (IIIb–IV)	Stratified by stage and analysed descriptively Study measured wait time to diagnosis and evaluated the cost of diagnosis and treatment based on stage of diagnosis. Assessed records for diagnostic testing in the 12 months prior to date of diagnosis Time to diagnosis from the initial diagnostic test was 5 months for stage IV patients vs. 6.2 months for stage I (mean ranged from 5 to 6.2 months for all stages). Multivariate regression analysis found that stage was the only significant independent predictor of total health-care cost ($p < 0.001$ for stage I vs. stage II, IIIa, IIIb, or IV)	No evidence of this
Gonzalez-Barcala 2014;[117] Spain	To analyse the delays in the diagnosis and treatment of lung cancer in our health area, the factors associated with the timeliness of care and their possible relationship with the survival of these patients	Patients with a cytohistologically confirmed diagnosis of lung cancer (2005–8) ($n = 307$)	• Retrospective cohort study (single centre) • [Medical records]	• Symptom to the first specialist consultation (T3) • First specialist consultation to confirmation of diagnosis (T11) • Confirmation of diagnosis to start of first treatment (T15) • First specialist consultation to start of treatment (T14)	Survival (3 years' follow-up)	Cox regression adjusted for sex, age, smoking status, stage, histology, comorbidity and general health status	No evidence of this

continued

TABLE 51 Systematic review 2: updated searches[a] (*continued*)

Author and year; country	Study aim	Patient population (n analysed)	Study design; [data sources]	Interval(s) definition (T interpretation), and how interval was analysed	Outcome measures	Comments and method of analysis	Addressed waiting time paradox?
				• Days; each interval divided into three or four categories with first used as reference: T3 and T14 ≥ 30, 31–60, 61–90, > 90 days; T11 and T15 ≥ 30, 31–60, > 60 days; T15			
Kim 2016;[119] Canada	Aim to quantify diagnostic and treatment intervals and identify factors associated with delays	Patients with histologically confirmed stage I–III NSCLC diagnosed and treated in Alberta (2004–11) ($n = 3009$)	• Retrospective cohort (multicentre) • [Cases identified from cancer registry, and data obtained from physician billing, inpatient/ outpatient hospital data and medical records]	• Initial chest X-ray (used as a proxy for the beginning of the diagnostic interval) to diagnostic biopsy (T13) • Diagnostic biopsy to first treatment (T15) • Initial chest X-ray to first treatment (T14) Days; 'delay' defined as: T14 ≥ 78, T13 ≥ 38, T15 ≥ 51 days; estimated the adjusted odds of being in the delay category for each stage [stage I, II, III; reference = stage I]	Stage	Multivariable logistic regression was carried out to identify factors associated with delays (including, among others, stage and histological type). Stage III was one of the factors associated with prompt care	Adjusted analyses, but unclear if adjusted for relevant features to wait time paradox
Radzikowska 2013;[115] Poland	To evaluate the influence on survival of delays in the diagnosis and treatment in an unselected population of SCLC patients	Patients admitted to pulmonary outpatients departments and registered on the National Tuberculosis and Lung Diseases Research Institute in Warsaw (1995–8). Survival data collected up until 2003 ($n = 3479$)	• Cohort study (unclear if prospective) (multicentre) • [Data on intervals and 5-year survival 'recorded using standard questionnaires']	'Patient's delay': first symptoms to first visit to the GP (T1) 'Doctor's delay'; first visit to GP to diagnosis (T8) Days; median used to divide into two groups: T1: ≤ 30 vs. > 30 days (shorter delays used as reference); T4: ≤ 42 vs. > 42 days (longer delay used as reference). [Note: 'Doctor's delay' was inconsistently defined as time from first doctor/GP visit to diagnosis (T8) and treatment (T9); marked here as T8]	Survival (5 years from diagnosis)	• Multivariate analysis included adjustment for clinical stage among other factors • Measured multiple intervals including time from first visit to doctor to first visit with specialist, first visit to specialist to diagnosis, diagnosis to treatment, symptom onset to diagnosis, and symptom onset to treatment but did not assess their impact on survival	No evidence of this

Author and year; country	Study aim	Patient population (n analysed)	Study design; [data sources]	Interval(s) definition (T interpretation), and how interval was analysed	Outcome measures	Comments and method of analysis	Addressed waiting time paradox?
Živković 2014;[116] Montenegro	To investigate whether or not waiting times and delays in diagnosis and treatment of patients with lung cancer have any bearing on prognosis and survival	Patients diagnosed and treated for lung cancer at a single hospital (2009) ($n = 206$)	• Retrospective cohort (single centre) • [Patient questionnaire and medical records]	• Symptoms' onset to first consultation with doctor (T1) • Onset of first symptoms to diagnosis (T4) • Weeks, divided into two groups: ≤ 8 vs. > 8 weeks to compare Kaplan–Meier curves	Survival (1 year, from symptom onset)	Survival analysed using Kaplan–Meier methods Assessed both 'patient delay' (T1) and 'total delay' (T4), which included patient and health-care delay (primary care 'doctor's delay I' and specialist medical services delay 'doctor's delay II') by comparing the cumulative survival for those seen within or after 8 weeks. There were no significant difference between the Kaplan–Meier curves for the assessment of either interval ($p > 0.05$)	No evidence of this
Lymphoma							
Nikonova 2015;[120] Canada	To describe the determinants of delays in diagnosis and treatment of patients with DLBCL and the impact of delays on clinical outcomes	Patients treated for DLBCL at a single cancer centre (2002–10) ($n = 278$)	• Retrospective cohort (single centre) • [Medical records, consensus data]	• First health-care contact to first haematologist contact: 'diagnostic wait time' (T7) (date of histopathology diagnosis categorised as occurring prior to or following initial haematologist contact) • First haematology visit to initiation of treatment: 'treatment wait time' (T14)	Survival (5-year OS/PFS rates)	• Kaplan–Meier method used to assess survival (Mantel–Cox log-rank test used for comparing survival curves for different intervals, and Cox regression used to assess whether or not delays are associated with survival) • Multivariate logistic regression used to assess association between clinically	No evidence of this

continued

TABLE 51 Systematic review 2: updated searches[a] (*continued*)

Author and year; country	Study aim	Patient population (n analysed)	Study design; [data sources]	Interval(s) definition (T interpretation), and how interval was analysed	Outcome measures	Comments and method of analysis	Addressed waiting time paradox?
				Weeks/months; categorised into three or four groups. T7: 0–1 week, > 1 week to 1 month, > 1 month; T14: 0–6 weeks, > 6 weeks to 3 months, 3–6 months, > 6 months. 'Delay' defined as: T7 > 6 months, T14 > 4 months • [Diagnostic delay was defined as a time to diagnosis of > 6 weeks, and treatment delay as time to treatment of > 4 weeks] • 14 patients (7%) had their diagnostic pathology after the initial haematologist visit		relevant covariates (inclining among other stages) and delays in diagnosis and treatment • Also define patient delay (T1), but do not report results	
Multiple sites							
Dregan 2013;[108] England, UK	To evaluate diagnostic time intervals and consultation patterns after presentation with alarm symptoms, and their association with cancer diagnosis and survival	• Patients presenting to the GP with a specific alarm symptom (haematuria, haemoptysis, dysphagia or rectal bleeding) or diagnosed with the corresponding cancer (urinary tract, lung, gastro-oesophageal, or colorectal, respectively) (2002–6) • ($n = 5524$: urinary tract, 842; lung, 215; gastro-oesophageal, 349; colorectal, 451)	• Retrospective population-based cohort study (multicentre) • [Primary care records linked with cancer registry]	First consultation with alarm symptom to diagnosis (T8) Days; divided into five groups: < 15, 15–90 (reference), 91–180, 181–365, and > 365 days	Survival (from diagnosis; up to 4 years' follow-up)	• Adjusted HR (95% CI) obtained from a Cox regression model (adjusted for age and sex) • Also analysed risk of death according to presence or absence of alarm symptoms	No evidence of this

Author and year; country	Study aim	Patient population (n analysed)	Study design; [data sources]	Interval(s) definition (T interpretation), and how interval was analysed	Outcome measures	Comments and method of analysis	Addressed waiting time paradox?
Redaniel 2015;[101] England, UK	To assess the association between diagnostic interval (time from presentation in primary care to diagnosis) and 5-year survival, stratified by NICE-qualifying alert and non-alert symptom presentations	Patients diagnosed with CRC, breast cancer, lung cancer or prostate cancer who presented to a GP with a cancer symptom 1 year prior to diagnosis identified (1998–2009) in England. (Excluded patients diagnosed through emergency routes) ($n = 22,051$: CRC, 5912; breast cancer, 8639; lung cancer, 5737; prostate cancer, 1763)	• Historical population-based cohort study (multicentre) • [Linked data sets: primary care records, cancer registry, deprivation indices, hospital statistics]	Presentation in primary care to diagnosis (T8) • Months for CRC, lung cancer and prostate cancer; divided into four groups: < 1 (reference), 1–2, 3–6, and > 6 months • Weeks for breast cancer; divided into four groups: < 1 (reference), 1–2, 3–4, and > 4 weeks	Survival (5 years)	EHR (95% CIs) computed using a generalised linear model with a Poisson error structure. Separate multivariable models built for each cancer site adjusting for, among others, the effects of period of cancer plan implementation, Dukes' stage, tumour subsite (for CRC) and tumour differentiation. Also included stratified analysis by alert vs. non-alert symptoms	No evidence of adjusting for symptoms
Myeloma							
Goldschmidt 2016;[121] Israel	To assess whether or not clinical and laboratory data could provide early clues to multiple myeloma diagnosis and whether or not time to detection affects survival	Patients diagnosed with multiple myeloma (2002–11), and matched cancer-free controls presenting with back pain, and who were insured in one of the Israeli health organisations (HMOs) ($n = 107$)	• Retrospective population-based study (case–control) (multicentre) • [Medical records and national cancer register]	First signs and symptoms of myeloma ('combination' of symptoms and laboratory results) to time of diagnosis (T8) Months; mean and categorised into three groups: < 2, 2–12 and > 12 months. Also compared Kaplan–Meier curves for the three groups	• Survival (from diagnosis) • Stage (higher)	• The impact of interval on survival rate assessed using Mann–Whitney, and interval groups compared using Kaplan–Meier curves (log-rank test). The impact of mean interval on stage assessed using Kruskal–Wallis test and interval categories using Fisher's exact test • Case–control design (and multivariate analysis) used to identify independent predictors of multiple myeloma • Included a period of 2 years before diagnosis	No evidence of this

continued

TABLE 51 Systematic review 2: updated searches[a] (*continued*)

Author and year; country	Study aim	Patient population (n analysed)	Study design; [data sources]	Interval(s) definition (T interpretation), and how interval was analysed	Outcome measures	Comments and method of analysis	Addressed waiting time paradox?
Neuroendocrine							
Keizer 2016;[122] the Netherlands	Note that the data extraction for this study is based on an abstract as the inter-library loan for the full text is still outstanding To investigate the association of delay in diagnosis with stage, grade, symptoms, primary site of the tumour and survival in patients with well-differentiated (grade 1, grade 2) neuroendocrine tumour	Patients referred to the NKI-AvL and Westfriesgasthuis (2003–9) ($n = 227$)	Unable to assess	Onset of first symptoms to diagnosis (T4) Months	• Survival • Stage	• Univariate linear regression analyses and Kaplan–Meier curves used to examine the association between delay in diagnosis, patient characteristics and survival • Patients with a longer delay were more likely to have psychological problems	Unable to assess
Oral							
Alahapperuma 2017;[123] Sri Lanka	To identify patient-linked delays between the time of first noticing symptoms and definitive diagnosis, and its association with stage at diagnosis of oral and pharyngeal carcinoma patients attending the National Cancer Institute	Patients with histologically confirmed carcinoma of the oral cavity or pharynx ($n = 351$)	• Cross-sectional study (single centre) • [Patient interview-questionnaire and data extraction sheet (diagnosis card, case notes, and referral letter)]	Symptoms' onset to first seen by health-care practitioner: 'patient-delay 1' (T1) Time taken to reach a specialty centre after being referred: 'patient-delay 2' (T10) Months/weeks; divided into two groups based on presence or absence of 'delay' defined as: > 3 months for T1, > 2 weeks for T10	Stage [III–IV (vs. I–II)]	• Possible associations were analysed using chi-squared test • 99 patients whose symptoms were first identified by health-care practitioner were excluded from analysis of T1; 44 patients who did not have a referral date were excluded from analysis of T10	No evidence of this

Author and year; country	Study aim	Patient population (n analysed)	Study design; [data sources]	Interval(s) definition (T interpretation), and how interval was analysed	Outcome measures	Comments and method of analysis	Addressed waiting time paradox?
Esmaelbeigi 2014;[124] Islamic Republic of Iran	To study the situation of professional delays and stage on diagnosis of oral cancer among patients who were hospitalised in the Cancer Institute hospital, the largest referral centre which treats head and neck cancer in the Islamic Republic of Iran	Oral cancer patients admitted to the Cancer Institute (2009–10) ($n = 206$)	• Retrospective cohort (single centre) • [Patient interviews] (Study was referred to as a case–control, where cases = patients with delay)	Onset of symptoms to initiation of treatment (T5) Days; stratified into two groups based on median, with higher than median (140 days) classified as 'delay'	Stage (advanced)	Multivariate logistic regression used to obtain OR (95% CI) adjusted for a number of patient characteristics (including symptoms) and grade Also assessed predictors of patient (onset of symptoms to first visit to physician) and professional (initial investigation symptoms to treatment) delay in diagnosis and treatment of oral cancer. Assessed association between 'total delay' (T5) and stage at diagnosis only	Unclear if adjustment for symptoms addresses this to any extent
Ovarian							
Altman, 2017;[126] Canada	To analyse data on time to diagnosis and correlate this with OS	All invasive epithelial ovarian cancer cases (diagnosed between 2004 and 2010) identified via the Manitoba Cancer Registry. Included emergency presentations and some patients whose cancer was identified incidentally (both presentation types adjusted for in analysis) ($n = 601$)	• Retrospective population-based cohort study (multisite) • [Medical records, CancerCare® charts, administrative data from Manitoba Health (Physician claims and hospital data)]	First presentation (contact with any health-care provider as a result of any symptoms related to epithelial ovarian cancer) to diagnosis (T8) Days; continuous variable	Survival (from diagnosis; at ≥ 4 years' follow-up)	• Secondary objectives were to investigate predictors of OS • Time-varying Cox regression analysis conducted, stratified by early (stage I–II, $n = 210$) and late stage (stage III–IV or unknown, $n = 391$). Analysis adjusted for various clinicopathological and socioeconomic factors, including stage at diagnosis, morphology, symptoms at first presentations, emergency presentation and cancer identified incidentally	Analysis adjusted for various stage, morphology, symptoms at first presentation, and emergency presentation

continued

TABLE 51 Systematic review 2: updated searches[a] (*continued*)

Author and year; country	Study aim	Patient population (n analysed)	Study design; [data sources]	Interval(s) definition (T interpretation), and how interval was analysed	Outcome measures	Comments and method of analysis	Addressed waiting time paradox?
						(e.g. asymptomatic and identified via screening or imaging for unrelated condition) ($n = 67$) • Restricted cubic splines were used for continuous predictors that violated the assumption of linearity (including time to diagnosis)	
Lim 2016;[125] UK	To compare time to diagnosis of the typically slow-growing type I (low-grade serous, low-grade endometrioid, mucinous, clear cell) and the more aggressive type II (high-grade serous, high-grade endometrioid, undifferentiated, carcinosarcoma) invasive epithelial ovarian cancer	Patients diagnosed with epithelial ovarian cancer (2006–8) recruited before definitive diagnosis or treatment. (Included emergency presentations.) ($n = 227$)	• Observational study (multicentre) • [Patient questionnaire, primary care records, hospital letters, and multidisciplinary team summaries]	First symptom to presentation (T1) First presentation to primary care to diagnosis (T8) Months; categorised into four: 0 to < 3, ≥ 3 to < 6, ≥ 6 to < 9, ≥ 9 months	Stage (advanced)	• Assessed relationship between patient/diagnostic intervals and FIGO stage for type I and type II cancers separately using the Wilcoxon rank-sum test • It was hypothesised that longer time to diagnosis would be associated with more advanced stage in type I invasive epithelial ovarian cancer, but not in type II invasive epithelial ovarian cancer (because of aggressive disease/rapid progression) • Included patients admitted via the emergency route with 222 (98%) patients completing questionnaires, of these, primary care records were available for 199 (88%) patients	No evidence of this

Author and year; country	Study aim	Patient population (n analysed)	Study design; [data sources]	Interval(s) definition (T interpretation), and how interval was analysed	Outcome measures	Comments and method of analysis	Addressed waiting time paradox?
Pancreatic							
Gobbi 2013;[127] Italy	To investigate the prognostic role of diagnostic delay and clinical presentation (regarding pain, jaundice, and weight loss) in pancreatic carcinoma	Newly diagnosed and treated patients with pancreatic cancer (2001–10) ($n = 170$)	• Retrospective cohort (single centre) • [Patient records]	Onset of the first symptom(s) to pathological diagnosis (T4) Weeks; continuous variable, also plotted separate Kaplan–Meier curves for three thresholds: < 4, 4–6 and > 16 weeks	Survival (cumulative; from diagnosis)	Kaplan–Meier methods (log-rank test). Multiple regression applied to survival according to the Cox proportional hazards model. Covariates included stage (stage I–II vs. stage III–IV), among others Diagnostic delay was one of the factors with a significant and independent prognostic value; stage was not. There was a significant different between the Kaplan–Meier curves for different thresholds ($p < 0.0001$), with longer delays associated with poor survival	No evidence of this
Jooste 2016;[128] France	To estimate patient and treatment delays in patients with pancreatic cancer and to measure their association with survival in a non-selected population	Patients diagnosed with pancreatic cancer for the first time (2009–11) and registered in two French digestive cancer registries. (Included emergency presentations.) ($n = 298$)	• Population-based cohort study (two regions) • [Cancer registries; interval data: GP questionnaire, medical and surgical records]	• 'Overall delay' (based on 'patient' plus 'treatment' delay): onset of symptoms to treatment (T5) • 'Patient delay': onset of symptoms to first consultation with practitioner (most often a GP) (T1) • 'Treatment delay': first practitioner consultation to treatment (T9)	Survival (3-year survival based on net survival)	• Univariate and multivariate analysis used to investigate covariates associated with patient and treatment delay • Survival analysis conducted for 'overall delay' only. 'Patient delay' analysed in 450 patients, 'treatment delay' in 345 and survival analyses performed on 298 patients for whom both delays were available	• Unclear if adjusted analyses for survival vs. delay included symptoms and type of first practitioner seen (emergency or otherwise) • Emergency presentations were included

continued

TABLE 51 Systematic review 2: updated searches[a] (*continued*)

Author and year; country	Study aim	Patient population (n analysed)	Study design; [data sources]	Interval(s) definition (T interpretation), and how interval was analysed	Outcome measures	Comments and method of analysis	Addressed waiting time paradox?
				Months/days; patient and treatment delays were categorised into two: T1 < 1 or ≥ 1 month; T9 < 29 or ≥ 29 days. Overall delay (T5) analysed as four categories based on the 'patient' and 'treatment' delay categories: 1. T1 < 1 month and T5 < 29 days (reference) 2. T1 < 1 month and T5 ≥ 29 days 3. T1 ≥ 1 month and T5 < 29 days 4. T1 ≥ 1 month and T5 ≥ 29 days		• Multivariate logistic regressions conducted, adjusting for, among other covariates, the presence of synchronous metastases (M0 vs. M+), and first practitioner (GP), emergency room, gastroenterologist, other specialist	
Penile							
Gao 2016;[129] China	To ascertain risk factors resulting in delayed treatment-seeking and to evaluate its influence on prognosis	Patients with a histopathologic diagnosis and treated for penile cancer (2004–10) ($n = 228$)	• Retrospective cohort (multicentre) • [Patient interview-questionnaire, medical records]	Symptom onset to first medical consultation (T1) Months; divided into four categories: ≤ 1 ['undelayed' reference], 1–3, 3–6, > 6 months	• Survival (OS) rate: ≥ 2 vs. < 2 years, and ≥ 5 vs. < 5 years • Stage • Size (> 2 cm) • Metastases (M1) • Lymph node involvement (N1–3)	• 254 patients enrolled in the study, but only 228 included in the survival analysis • Multivariate logistic regression used to assess impact of different intervals on OS and lesion characteristics, with OR (95% CI) (Fisher's exact test) estimated for each 'delay' interval, compared with 'undelayed'. Multivariate analysis also used to investigate risk factors associated with patient delay. Lesion characteristics included size, stage, lymph node involvement and distant metastases	No evidence of this

Author and year; country	Study aim	Patient population (n analysed)	Study design; [data sources]	Interval(s) definition (T interpretation), and how interval was analysed	Outcome measures	Comments and method of analysis	Addressed waiting time paradox?
Prostate							
Bonfill 2015;[130] Spain	To determine the clinical characteristics of prostate cancer patients, the diagnostic process and the factors that might influence intervals from consultation to diagnosis and from diagnosis to treatment	Symptomatic and asymptomatic patients diagnosed and treated for prostate cancer at one of seven tertiary care hospitals (2010–11) ($n = 470$)	• Retrospective cohort study (multicentre) • [Medical records and patient inteviews]	• First presentation to a health-care professional to confirmed diagnosis (T8) • First presentation to a health-care professional to confirmed treatments (T15) Days; dichotomised: T8 ≤ 100 vs. > 100 days; T15 ≤ 30 vs. > 30 days	Stage (T2a–T4; reference: T1a–T1c)	Multivariate analysis also used to investigate association between intervals and diagnostic characteristics (potential predictors); no significant association between the interval groups and stage found in univariate analysis for T8 or T15. EMPARO-CU study	Unlikely. Measured number of symptoms, but results based on univariate analyses
Sarcoma							
De Boer 2014;[133] Uganda	To measure the association between primary delay and the cancer stage of HIV Kaposi's sarcoma patients on diagnosis of the disease by a clinician	HIV-infected adults with histologically-confirmed Kaposi's sarcoma treated at a cancer institute (2012) ($n = 161$)	• Cross-sectional study (single centre) • [Interviews for intervals, and medical records for stage at admission]	Symptom onset to presentation to a clinician (any health professional) (T1) Months; dichotomised: ≤ 3 vs. > 3 months	Stage (poor-risk stage)	Multivariable logistic regression used to estimate OR of having poor-risk stage vs. low-risk stage at presentation, adjusted for age, sex, income, and exposure to antiviral therapy. Also evaluated factors associated with delay	No evidence. Measured pain. But it was not included in multivariate analysis

continued

TABLE 51 Systematic review 2: updated searches[a] (*continued*)

Author and year; country	Study aim	Patient population (n analysed)	Study design; [data sources]	Interval(s) definition (T interpretation), and how interval was analysed	Outcome measures	Comments and method of analysis	Addressed waiting time paradox?
Goedhart 2016;[132] the Netherlands	To investigate delay in diagnosis by both patients and doctors, and to evaluate its effect on outcomes of high-grade sarcoma of bone	Patients with high-grade bone sarcoma diagnosed (2000–12) at a single oncological centre ($n = 102$)	• Retrospective cohort (single centre) • [GP records (GPs asked for date of symptom onset) and hospital]	'Total delay': initial symptoms to histopathological diagnosis (T4) 'Referral clinic doctor-related delay': presentation to oncology centre and histopathological diagnosis (T11) Months/days; dichotomised: T4 < 4 vs. ≥ 4 months; T11 ≤ 42 vs. > 42 days	• Survival (5-year OS rate) • Metastases	• Kaplan–Meier method • Measured various different intervals to assess delays, but did not assess impact of each of these on survival	No evidence of this
Urakawa 2015;[131] Japan	To determine whether or not the duration from initial symptoms to specialist consultation affects the prognosis for soft-tissue sarcoma	Patients with primary soft-tissue sarcoma seen at a single specialist hospital (2001–11). Excluded patients seen within 1 month (most were second-opinion patients) ($n = 142$)	• Retrospective cohort (single centre) • [Patient records and hospital database]	Onset of symptoms to specialist consultation (T3) Months; divided into two groups: < 6 vs. ≥ 6 months (reference); also plotted separate Kaplan–Meier curves	Survival (5-year OS rate; from first seen by specialist)	Kaplan–Meier methods (log-rank test). Cox proportional hazard methods adjusting for multiple covariates, including, among others, tumour size, depth, and histological grade used to obtain HR (95% CIs) Data on delay available for 142 patients with soft-tissue sarcoma, of whom 142 had non-small cell sarcoma and were included in survival analysis	Excluded patients seen within 1 month (most were second-opinion patients)

Author and year; country	Study aim	Patient population (n analysed)	Study design; [data sources]	Interval(s) definition (T interpretation), and how interval was analysed	Outcome measures	Comments and method of analysis	Addressed waiting time paradox?
Testicular							
Kobayashi 2014;[134] Japan	To clarify the effect of the time to consultation on disease extension and survival	Patients who were treated for testicular germ cell tumours at a cancer centre (1991–2010) ($n = 175$)	• Retrospective cohort (single centre) • [Cancer registry and medical records]	Symptom onset to first medical consultation at cancer centre or associated institution (no appointment needed, so system issues were absent) (T1) Months; divided into two groups: < 6 vs. ≥ 6 months (reference); also plotted separate Kaplan–Meier curves	• Survival (5-year OS rate, from diagnosis) • Stage (I vs. II–III) • Size (mean and SD)	• Kaplan–Meier method (log-rank test). Multivariate analysis adjusting for histology and stage). Multivariate analyses (Cox proportional hazard methods) adjusting for histology and stage used to obtain HR (95% CIs). Differences between interval groups also compared for stage (chi-squared test) and tumour size (Welch-correct *t*-test, and Kruskal–Wallis test) • Also evaluated shift in consultation time the last two decades • Disease stages were similar between the two interval groups ($p = 0.897$)	No evidence of this

DLBCL, diffuse large B-cell lymphoma; EHR, excess hazard ratio; FIGO, Fédération Internationale de Gynécologie et d'Obstétrique (International Federation of Gynaecology and Obstetrics); HIV, human immunodeficiency virus; HMO, Health Maintenance Organization; M, metastasis; N, node; NKI-AvL, Nederlands Kanker Instituut Antoni van Leeuwenhoek; NSCLC, non-small-cell lung cancer; OR, odds ratio; OS, overall survival; PFS, progression-free survival; SCLC, small-cell lung cancer; SD, standard deviation.

a Ordered alphabetically by cancer site and then author, with studies evaluating more than one cancer reported under 'multiple'.

TABLE 52 Description of studies included in the updated review

Study	Outcome	Time interval[a]	Association	Outcome description	Statistic test/method of analysis	Results
Aslam 2017[110]	Survival	T12	<>	Median OS	Log-rank test comparing median survival from Kaplan–Meier curves. The study also included patients referred via screening, survival analysis limited to emergency, 2WW, urgent and routine referrals. Analysis also conducted excluding emergency	There was no significant difference in overall survival between the two interval groups (for 2WW, urgent, and routine referrals combined): < 62 days group: 7.1 years vs. > 62 days group: 6.54 years ($p = 0.620$)

Chen 2017[112]	Stage	T4
		T1
		T8
			– Aged < 50 years
			– Aged < 50 years
			<>
				TNM stage: advanced (stage III–IV) vs. non-advanced (stage I–II)	Median interval (days) from symptom onset and/or workup to diagnosis were compared for advanced and non-advanced stage at diagnosis (but not statistically) within each age group	The authors noted that the longest time to diagnosis (T4) was observed in patients aged < 50 years with non-advanced CRC at diagnosis (stage I and II) followed by patients aged < 50 years with advanced CRC at diagnosis (stages III and IV) (median days 174 vs. 124)
						• Aged < 50 years, non-advanced vs. advanced stage
						  ○ Median symptom duration (T1): 90 vs. 60 days
						  ○ Median workup duration (T8): 39 vs. 29 days
						• Aged ≥ 50 years, non-advanced vs. advanced stage:
						  ○ Median symptom duration (T1): 21 vs. 30 days
						  ○ Median workup duration (T8): 31 vs. 17 days

Study	Outcome	Time interval[a]	Association	Outcome description	Statistic test/method of analysis	Results
Dregan 2013[108]	Survival	T4	<>	Overall mortality (from diagnosis)	Adjusted HR (95% CI) obtained from Cox regression model (adjusted for age and sex). (Also analysed risk of death according to presence or absence of alarm symptoms)	• ≤ 14 days ($n = 79$) 1.07 (95% CI 0.65 to 1.77); $p = 0.794$ • 15–90 days ($n = 217$), reference category • 91–180 days ($n = 56$) 1.28 (95% CI 0.73 to 2.25); $p = 0.386$ • 181–365 days ($n = 44$) 0.76 (95% CI 0.38 to 1.54); $p = 0.448$ • > 365 days ($n = 55$) 0.92 (95% CI 0.45 to 1.85); $p = 0.806$ Patients with no preceding alarm symptoms had shorter survival from diagnosis than those presenting with relevant alarm symptoms
Helewa 2013[106]	Survival	T9	<>	5-year OS	Adjusted HR (95% CI) and *p*-values obtained from multivariate Cox proportional hazards model using bootstrap resampling. Important confounders adjusted for included, among others, stage, emergency presentations and location. Emergency presentations were also analysed separately. Also conducted analysis limited to stage I–III CRC (excluding stage IV and 'unknown')	Increased total wait time quartile was not associated with worse survival ($p = 0.4898$) • Q1 0 to ≤ 43 days ($n = 175$): reference • Q2 ($n = 135$) > 43 to ≤ 95 days: 0.93 (95% CI 0.73 to 1.19) • Q3 ($n = 115$) > 95 to ≤ 166 days: 0.90 (95% CI 0.69 to 1.16) • Q4 ($n = 113$) > 166 to ≤ 513 days: 0.82 (95% CI 0.64 to 1.06) When total wait time was considered as a continuous variable (days), it was still not associated with worse survival ($p = 0.2618$)
Janssen 2016[114]	Stage (advanced)	T11	<>	Node positivity, and presence of distant metastases	Chi-squared tests	• Node positive in < 60 days group: 42/102 (41%) vs. > 60 days group: 52/144 (36%) ($p = 0.42$) • Distant metastases in < 60 days group: 12/102 (12%) vs. > 60 days group: 9/144 (6%) ($p = 0.13$)

continued

TABLE 52 Description of studies included in the updated review (*continued*)

Study	Outcome	Time interval[a]	Association	Outcome description	Statistic test/method of analysis	Results
Leiva 2017[113]	Stage (advanced)	T8	• – (Hospital) • <> (GP)	TNM tumour stage (0, 1, 2, 3, 4, unknown)	The Mann–Whitney *U*-test and the Kruskal–Wallis test were used to assess association between diagnostic intervals and various patient and clinical characterises (including, among others stage, emergency presentation, location, symptoms at presentation, and perceived seriousness of symptoms). The independent effect of variables was assessed using Cox regression and then the extension of the multivariate Cox proportional hazards model with time-dependent covariates	• Comparing three different information sources for estimating 'diagnostic intervals': patient recall, hospital records and GP records. Interval defined as date from first registry of symptoms to date of diagnosis for GP and hospital records (T8), but, for patient recall, this was time from patient first experiencing symptoms to date of diagnosis (T4) • Univariate analysis indicated that a shorter diagnostic interval was associated with more advanced tumour stage using both GP and hospital records, but this was significant for hospital records data only ($p = 0.021$) • Stage was not an independent predictor of a shorter interval in the multivariate analysis of hospital and GP data
		T4	<> (patient)	TNM tumour stage (0, 1, 2, 3, 4, unknown)	As above	• Univariate analysis indicated that a shorter diagnostic interval was associated with more advanced tumour stage for patient recall, but this was not significant • Stage was identified as one of the independent predictors of shorter interval in the multivariate analysis for patient-recorded data (but not in the analysis of hospital and GP data)

Study	Outcome	Time interval[a]	Association	Outcome description	Statistic test/method of analysis	Results
Murchie 2014[107]	Stage (advanced)	T9	<>	Dukes' stage was collapsed into a binary variable: early (A or B) vs. advanced (C or D)	OR (99% CI) obtained from logistic regression model with a restricted cubic spline (to allow for non-linear relationship), following sequential adjustment (using 1–4 models; 1 representing unadjusted) of patient and tumour factors (including, among others, tumour grade, and signs and symptoms). Time to treatment was modelled using four knots corresponding to the 25th, 50th, 75th and 100th centile	• The spline curves showed a significant non-linear ($p = 0.04$) (and unadjusted, $p = 0.002$) association between provider delay and stage. Delays of between 4 and 34 weeks were associated with earlier stage, and intervals beyond this with later stage disease. However, the plot based on a fully adjusted model showed wide CIs • According to the fully adjusted model, provider delays of 40 and 60 weeks were associated with later-stage disease at presentation, but the OR did not reach statistical significance o 4 weeks (reference) o 20 weeks 0.87 (99% CI 0.54 to 1.39) o 40 weeks 1.46 (99% CI 0.93 to 2.31) o 60 weeks 1.58 (99% CI 0.96 to 2.61)
Murchie 2014[107]	Survival	T9	<>	All-cause survival (from date of first presentation)	HR (99% CI) obtained from Cox survival models, both with a restricted cubic spline (to allow for non-linear relationship), following sequential adjustment (1–4 models) of patient and tumour factors (including, among others, tumour grade, emergency admission, and signs and symptoms). Time to treatment was modelled using four knots corresponding to the 25th, 50th, 75th and 100th centile. Stratified analysis also conducted for rectal and colon cancer	The spline curves showed a significant non-linear ($p < 0.001$) (and unadjusted inverse, $p < 0.001$) association between provider delay and mortality. In the univariate analysis, diagnostic interval of < 4 weeks was associated with poor survival, but this was no longer present after adjusting for confounders. According to the fully adjusted model, provider delays of 40 and 60 weeks were associated with later-stage disease at presentation, but the OR did not reach statistical significance • 4 weeks (reference) • 20 weeks 0.99 (99% CI 0.76 to 1.27) • 40 weeks 1.17 (99% CI 0.92 to 1.48) • 60 weeks 1.21 (99% CI 0.94 to 1.57)

continued

TABLE 52 Description of studies included in the updated review (*continued*)

Study	Outcome	Time interval[a]	Association	Outcome description	Statistic test/method of analysis	Results
Patel 2018[111]	Survival	T12	–	Long-term (mean) survival	Proportion of patients still alive at long-term follow-up (October 2011), and log-rank test comparing Kaplan–Meier curves. Also assessed 30-day postoperative mortality and cause of death	• Long-term survival was greatest in elective patients who did not receive treatment > 62 days (52% alive), compared with elective cases treated ≤ 62 days (34% alive). This was supported by the log-rank analysis ($p < 0.001$) • Operative mortality was higher in patients treated ≤ 62 days (7% elective, 20% emergency) than in those who were treated > 62 days (4% elective, 7% emergency). The most common cause of death was CRC in all groups
Patel 2018[111]	Stage	T12	–	Stage: Dukes' A–B vs. C–D	Proportions compared using chi-squared test	• Proportion of early-stage disease (Dukes' stage A–B) was highest in elective patients treated > 62 days (50%) ($p < 0.01$) and lowest in emergencies treated ≤ 62 days (30%) ($p = 0.26$) • Later-stage disease (Dukes' stage C–D) was most common in emergency patients treated ≤ 62 days (58%) and lowest in elective patients treated > 62 days (36%)

Study	Outcome	Time interval[a]	Association	Outcome description	Statistic test/method of analysis	Results
Pita-Fernández 2016[109]	Survival	T4	• <> (rectal, not adjusted for stage) • <> (colon) • <> CRC	5-year mortality (after diagnosis)	The influence survival was analysed in two ways: 1. Kaplan–Meier survival curves were computed for each interval quartile, and compared using the log-rank test 2. the interval was treated as a continuous variable using restricted cubic splines with four knots and using the 50th percentile (3.4 months) as reference point The 5-year HRs (95% CI) were estimated as a function of the length of the delay interval and adjusted for age and sex (model 1), and subsequently stage (model 2), using proportional hazard Cox regression	• The Cox regression model adjusting for age and sex showed that, in rectum cancers, patients within the first interval quartile had lower survival ($p = 0.003$), but this was no longer statically significant when also adjusting for stage ($p = 0.084$) • No significant differences were found for colon cancer when adjusting for age and sex ($p = 0.282$) or also adjusting for stage ($p = 0.160$) • Longer intervals were not associated with poor survival in CRC patients. From age-, sex- and stage-adjusted model: o < 1.5 months (reference) o 1.5–3.4 months, 0.97 (95% CI 0.71 to 1.32) o 3.4–6.4 months, 0.76 (95% CI 0.55 to 1.04) o > 6.4 months, 0.77 (95% CI 0.56 to 1.05) • The cubic splines regression analysis revealed that, for rectum cancer, 5-year mortality progressively increases for intervals less than the median (3 months) and decreases as the delay increases until approximately 8 months. In colon cancer, no significant relationship was found between interval and survival
Pita-Fernández 2016[109]	Stage (advanced)	T4	<>	TNM tumour stage (I, II–III, IV)	The association between the interval (categorised into quartiles) and tumour stage analysed using multivariate logistic regression (adjusting for age and sex)	The interval was not found to be significantly associated with stage at diagnosis [stage I: 3.6 months; stage II–III: 3.4 months; stage IV: 3.2 months ($p = 0.728$)], even after adjusting for age and sex

continued

TABLE 52 Description of studies included in the updated review (*continued*)

Study	Outcome	Time interval[a]	Association	Outcome description	Statistic test/method of analysis	Results
Pruitt 2013[137]	Survival	T8	+ (Colon) <> (Rectal)	All-cause mortality, and CRC-specific mortality	Cases (CRC deaths) and controls (deaths due to other causes or censored) were matched on survival time and the association of delay with death was examined using logistic regression. OR (95% CI) estimated using multivariate regression analysis adjusting for sociodemographic, tumour (histology, stage, grade, location), and treatment factors. Stratified analysis also conducted according to and stage (and presenting symptom in sensitivity analysis)	• Colon cancer patients with the longest diagnostic delays (8–12 months vs. 14–59 days) had higher odds of all-cause (OR 1.31, 95% CI 1.08 to 1.58), but not CRC-specific, death • Among rectal cancer patients, delays were not associated with risk of all-cause or CRC-specific death
Pruitt 2013[137]	Survival	T15	• – (Colon, < 1 week) • <> (Rectal)	All-cause mortality, and CRC-specific mortality	OR (95% CI) estimated using multivariate regression analysis adjusting for sociodemographic, tumour (histology, stage, grade, location), and treatment factors. Stratified analysis also conducted according to and stage (and presenting symptom in sensitivity analysis)	• Colon cancer patients with the shortest treatment delays (< 1 vs. 1–2 weeks) had higher odds of all-cause (OR 1.23, 95% CI 1.01 to 1.49), but not CRC-specific, death • Among rectal cancer patients, delays were not associated with risk of all-cause or CRC-specific death
Redaniel 2015[101]	Survival	T8	<>	5-year survival	5-year EHRs (95% CI) were computed using a generalised linear model with a Poisson error structure. Univariable and multivariable models built, for each cancer site, and stratified by the nature of the symptoms (NICE-qualifying alert vs. non-alert). Multivariable models controlled for, among others, the effects of period of cancer plan implementation, Dukes' stage, tumour subsite (for CRC), and tumour differentiation	There was no evidence of an association between diagnostic interval and mortality. EHR based on multivariate analysis for all patients: • < 1 month (reference) • 1–2 months 0.94 (95% CI 0.82 to 1.07) • 3–6 months 0.92 (95% CI 0.82 to 1.04) • > 6 months 0.93 (95% CI 0.81 to 1.07) The results showed significant lower mortality associated with longer diagnostic intervals for patients presenting with non-alert symptoms (EHR > 6 months vs. < 1 month: 0.85, 0.72 to 1.00; $p = 0.049$), whereas no significant associations were identified for patients presenting with NICE-qualifying alert symptoms

Study	Outcome	Time interval[a]	Association	Outcome description	Statistic test/method of analysis	Results

Tørring 2013[99]	Survival	T8	– (Alarm symptoms, Q1); <> (Vague symptoms)	5-year overall mortality	• Analyses stratified according to the GP's interpretation of the presenting symptoms: alarm or serious vs. vague. OR (95% CI) obtained from logistic regression using restricted cubic splines and adjusting for comorbidity, age, and sex. Each model tested against a model with no diagnostic interval term using the Wald test.
					• Data on tumour stage and emergency admission collected, and used to describe patient characteristics and compare the two groups, with alarm/serious symptoms or vague symptoms
					
					(Note: this appears to be the same CRC cohort as Tørring, 2011[136])
						• Alarm/serious symptoms ($n = 201$): the categorical analysis showed that both very short and long diagnostic intervals (first or fourth quartile compared with second and third) were associated with poor survival, but only the adjusted OR for the former was statistically significant:
						  ○ Q1: 4.74 (95% CI 2.20 to 10.19)
						  ○ Q2 + Q3: reference
						  ○ Q4: 2.01 (95% CI 0.93 to 4.36)
						
						• This association was confirmed by the spline regression analyses, which revealed convex (U-shaped) associations: the risk of dying within 5 years decreased with longer diagnostic intervals up to approximately the 60th percentile and then increased ($p = 0.001$)
						• Vague symptoms ($n = 67$): the categorical analysis showed that the first diagnostic interval quartile was associated with poor survival, compared with the second and third, whereas the fourth was not, but these findings were not statistically significant. Adjusted OR for each quartile:
						  ○ Q1: 0.74 (95% CI 0.19 to 2.80)
						  ○ Q2 + Q3: reference
						  ○ Q4: 1.08 (95% CI 0.28 to 4.12)
						
						• The corresponding spline regression analyses did not show a significant association ($p = 0.620$)

continued

TABLE 52 Description of studies included in the updated review (*continued*)

Study	Outcome	Time interval[a]	Association	Outcome description	Statistic test/method of analysis	Results
Tørring 2012[135]	Survival	T8	−/+ (GP-based; Q1, Q4) <> (Patient-based; GP records-based) −/+ (combined; Q1, Q2)	5-year overall mortality rate (after diagnosis)	Data taken from three studies using different methods for identifying date of first presentation. Data analysed for each study separately and combined The data were modelled in two ways: 1. Comparison of the first and fourth diagnostic interval quartiles with the second and third 2. Diagnostic interval rescaled and treated as a continuous variable using restricted cubic splines with four knots with the 50th percentile used as the reference point The 5-year HR (95% CI) estimated as a function of the length of the diagnostic interval and adjusted for tumour site (colon/rectum), age and sex using proportional hazard Cox regression. In the analysis of the combined data, differences were also allowed for in study-specific baseline hazards. (Note that the GP record-based study is Tørring 2011[136])	GP-based study ($n = 266$) adjusted HRs for each quartile: • Q1: 1.73 (95% CI 1.16 to 2.59) • Q2 + Q3: reference • Q4: 1.75 (95% CI 1.17 to 2.63) Patient-based study ($n = 658$) adjusted HRs for each quartile: • Q1: 1.26 (95% CI 0.95 to 1.66) • Q2 + Q3: reference Q4: 1.11 (95% CI 0.85 to 1.46) Record-based study ($n = 319$) adjusted HRs for each quartile: • Q1: 1.20 (95% CI 0.84 to 1.71) • Q2 + Q3: reference Q4: 1.30 (95% CI 0.91 to 1.86) Combined data ($n = 1243$) adjusted HRs for each quartile: • Q1: 1.33 (95% CI 1.10 to 1.61) • Q2 + Q3: reference Q4: 1.28 (95% CI 1.06 to 1.55) The association between diagnostic interval and survival was the same for all three types of data: displaying a U-shaped association with decreasing and, subsequently, increasing mortality with longer diagnostic intervals (i.e. U-shaped association observed using three different data collection methods in different health-care systems and over different time periods)

Study	Outcome	Time interval[a]	Association	Outcome description	Statistic test/method of analysis	Results
Tørring 2011[136]	Survival	T8	– (Alarm symptoms) <> (Vague symptoms)	3-year overall mortality (after diagnosis)	• Analyses stratified according to the GP's interpretation of the presenting symptoms: alarm or serious vs. vague. Logistic regression used to estimate 3-year mortality ORs (95% CIs) as a function of the diagnostic interval using restricted cubic splines and adjusting for tumour site, comorbidity, age and sex. Also reported were outcome data for 1-year mortality • Data on tumour stage and emergency admission collected, and used to describe patient characteristics and compare the two groups, with alarm/serious symptoms or vague symptoms	• Alarm/serious symptoms ($n = 201$): the cubic spline regression analysis revealed that the risk of dying within 3 years decreased with diagnostic intervals of up to 5 weeks and then increased ($p = 0.002$) • Adjusted OR for diagnostic intervals: o 0–4 weeks ($n = 75$), 2.56 (95% CI 1.29 to 5.05) o 5–11 weeks ($n = 90$): reference o ≥ 12 weeks ($n = 36$): 2.04 (95% CI 0.87 to 4.77) • Vague symptoms ($n = 67$): the cubic spline regression analysis revealed a reverse effect, with increasing risk of dying from day 1 up to ≈ 12 weeks, but the association was not statistically significant ($p = 0.205$) • Adjusted OR for diagnostic intervals: o 0–4 weeks ($n = 10$): comparison not justified o 5–11 weeks ($n = 27$): reference o ≥ 12 weeks ($n = 30$): 0.71 (95% CI 0.32 to 2.91)

+, positive association, based on a statistically significant more favourable outcome; –, negative association, based on a statistically significant less favourable outcome; <>, findings that were not statistically significant; EHR, excess hazard ratio; HR, hazard ratio; OS, overall survival; Q, quartile.
a See *Figure 2* for the time intervals.

Results of the studies included in the updated review

The 'abbreviated results' presented in *Table 53* represent those used to populate *Table 11*. The results provide a summary of the impact of 'short intervals' on outcome to correspond to the methods used in the original review.

These abbreviated results, however, represent a very simplistic interpretation of the study findings. Some studies assessed the interval as a categorical variable, whereby not all categories were associated with a significant or the same findings. Some studies evaluated the impact of 'long intervals' (often referred to as 'delay'), which were converted to represent the impact of 'short intervals'. A brief summary of the study findings are presented in the last column of *Table 53* to show how our 'abbreviated results' were derived from the actual study results for these studies. It was not our intention to extract detailed results of included studies for non-CRC sites; therefore, the brief summaries relate mainly to the interpretation of studies reporting statistically significant findings.

TABLE 53 Updated review: results of included studies[a]

Study	Outcome	Time interval[b]	Abbreviated results	Summary of results for studies reporting significant findings
Breast				
Ángeles-Llerenas 2016[100]	Survival	T5	<>	
Ángeles-Llerenas 2016[100]	Survival	T1	<>	
Ángeles-Llerenas 2016[100]	Survival	T15	<>	
Murchie 2015[102]	Survival	T9	–	Very short intervals – poor outcomes • Patients with very short intervals (up to 4 weeks) had poor prognosis, but longer delays were not associated with worse survival
Murchie 2015[102]	Stage (III–IV)	T9	–	Very short intervals – poor outcomes • Patients with very short intervals (up to 4 weeks) had poor prognosis, but longer delays were not associated with more advanced stage

Pace 2015[103]	Stage (advanced)	T1	+	Longer intervals – worse outcomes
				• Intervals of ≥ 6 to < 12 and ≥ 6 months only were significantly associated with more advanced-stage disease (compared with < 3 months)
				• OR and 95% CI for the three interval comparisons presented as a forest plot only
				  ○ ≥ 3 to < 6 vs. < 3 months: not significant
				  ○ ≥ 6 to 12 vs. < 3 months: OR > 3 months with CI not crossing 1
				  ○ > 12 vs. < 3 months: OR > 3 months with CI not crossing 1

TABLE 53 Updated review: results of included studies[a] (*continued*)

Study	Outcome	Time interval[b]	Abbreviated results	Summary of results for studies reporting significant findings
Pace 2015[103]	Stage (advanced)	T8	+	Longer intervals – worse outcomes • Intervals of ≥ 6 to < 12 months were significantly associated with more advanced-stage disease (compared with < 3 months) • OR and 95% CI for the three interval comparisons presented as a forest plot only o ≥ 3 to < 6 vs. < 3 months: not significant o ≥ 6 to 12 vs. < 3 months: OR > 3 months with CI not crossing 1 o > 12 vs. < 3 months: not significant
Unger-Saldaña 2015[104]	Stage (advanced)	T3	+	Longer intervals – worse outcomes • For every increase of 1 month in the 'patient interval' (T3), there was a 1.8% ($p < 0.01$) rise in the probability of beginning cancer treatment in advanced stages (stage III–IV vs. stage 0, I or II)
Unger-Saldaña 2015[104]	Stage (advanced)	T14	+	Longer intervals – worse outcomes • For each increment of 1 month in the 'provider interval' (T14), the probability of beginning cancer treatment in advanced stages (stage III–IV vs. stage 0, I or II) increased by 1% ($p < 0.05$)
Bladder				
Bryan 2015[105]	Survival	T5	<>	
Bryan 2015[105]	Survival	T12	<>	
Bryan 2015[105]	Survival	T2	<>	
Bryan 2015[105]	Survival	T10	<>	
Bryan 2015[105]	Survival	T14	<>	
Bryan 2015[105]	Stage (advanced)	T5	<>	
Bryan 2015[105]	Stage (advanced)	T12	<>	
Bryan 2015[105]	Stage (advanced)	T2	<>	
Bryan 2015[105]	Stage (advanced)	T10	<>	
Bryan 2015[105]	Stage (advanced)	T14	<>	
Bryan 2015[105]	Tumour size (> 2 cm)	T14	–	Longer intervals – better outcomes • Longer delays (> 20 days) in T3 were associated with smaller tumour size ($p < 0.05$) • Number (%) of patients in each interval group for tumour size of ≤ 2 cm and of > 2 cm: o ≤ 2 cm: > 20 days, $n = 244$ (37%) vs. ≤ 20 days, $n = 302$ (43%) o > 2 cm: > 20 days, $n = 412$ (63%) vs. ≤ 20 days, $n = 395$ (57%)

continued

TABLE 53 Updated review: results of included studies[a] (*continued*)

Study	Outcome	Time interval[b]	Abbreviated results	Summary of results for studies reporting significant findings
Bryan 2015[105]	Tumour size (> 2 cm)	T12	–	Longer intervals – better outcomes • Longer delays (> 68 vs. ≤ 68 days) in 'Hospital Delay' were associated with smaller tumour size ($p < 0.05$) • Number (%) of patients in each interval group for tumour size of ≤ 2 cm and of > 2 cm: o ≤ 2 cm: > 20 days, $n = 247$ (36%) vs. ≤ 20 days, $n = 303$ (45%) o > 2 cm: > 20 days, $n = 439$ (64%) vs. ≤ 20 days, $n = 379$ (55%)
Colorectal				
Aslam 2017[110]	Survival	T12	<>	
• Chen 2017[112]	• Stage (III–IV)	• T4	– (aged < 50 years)	Shorter intervals – worse outcomes *Among younger patients, those with advanced stage (stage III–IV) had shorter symptom and workup durations than early stage (stage I-II) CRC* • [No statistical analysis conducted, but the difference in median interval for patients with advanced vs. non-advanced stage (for young-onset CRC) was 40 days]
• Chen 2017[112]	• Stage (III–IV)	• T1	– (aged < 50 years)	Shorter intervals – worse outcomes *Patients with advanced stage had shorter symptom durations than early stage* • No statistical analysis conducted, but for young-onset CRC, the difference in median 'symptom duration' for patients with advanced and non-advanced stage was 30 days
Chen 2017[112]	Stage (III–IV)	T8	<> (aged < 50/≥ 50 years)	
Chen 2017[112]	Stage (III–IV)	T4	<> (aged ≥ 50 years)	
Chen 2017[112]	Stage (III–IV)	T1	<> (aged ≥ 50 years)	
Helewa 2013[106]	Survival	T9	<>	
Janssen 2016[114]	Node positivity, and presence of distant metastases	T11	<>	
Leiva 2017[113]	Stage (more advanced)	T8	– (hospital) <> (GP)	Longer intervals – better outcomes • The diagnostic interval obtained from hospital records was greater in those with less advanced CRC ($p = 0.021$)

TABLE 53 Updated review: results of included studies[a] (*continued*)

Study	Outcome	Time interval[b]	Abbreviated results	Summary of results for studies reporting significant findings
Leiva 2017[113]l	Stage (more advanced)	T4	<>	
Murchie 2014[107]	Stage (advanced; C–D)	T9	<>	
Murchie 2014[107]	Survival	T9	<>	
Pita-Fernández 2016[109]	Survival	T4	<>	
Patel 2018[111]	Survival	T12	–	Longer intervals – better outcomes • Long-term survival was greatest in elective patients who failed the standard, with 52% alive in October 2011, compared with 34% of elective cases meeting the standard ($p < 0.001$)
Patel 2018[111]	Stage	T12	–	Longer intervals – better outcomes • The proportion of early-stage disease (Dukes' stage A and B) was highest in elective patients who failed the standard (50%; $p < 0.01$) and lowest in emergencies meeting the standard (30%; $p = 0.26$)
Multisite				
Urinary tract				
Dregan 2013[108]	Survival	T8	+	Longer intervals – worse outcomes • Longer diagnostic intervals were associated with increased mortality (HR 2.23, 95% CI 1.35 to 3.69)
Lung				
Dregan 2013[108]	Survival	T8	<>	
Gastro-oesophageal				
Dregan 2013[108]	Survival	T8	<>	
Colorectal				
Dregan 2013[108]	Survival	T8	<>	
Prostate				

Redaniel 2015[101]	Survival	T8	–	Longer intervals – better outcomes
				Prostate cancer mortality was lower in patients with longer diagnostic intervals, regardless of type of presenting symptom
				• Outcome = mortality; interval in months
				  ○ 1–2 months vs. < 1 month: EHR 0.64, 95% CI 0.42 to 0.98
				  ○ > 6 months vs. < 1 month: EHR 0.45, 95% CI 0.27 to 0.75
				  ○ (No significant difference for 3–6 months vs. < 1 month: 0.72, 95% CI 0.49 to 1.05)

continued

TABLE 53 Updated review: results of included studies[a] (*continued*)

Study	Outcome	Time interval[b]	Abbreviated results	Summary of results for studies reporting significant findings
Lung				

Redaniel 2015[101]	Survival	T8	–	Longer intervals – better outcomes
				• Interval – months
				  ○ 3–6 months vs. < 1 month: EHR 0.90, 95% CI 0.84 to 0.98
				  ○ > 6 months vs. < 1 month: EHR 0.87, 95% CI 0.80 to 0.94
				  ○ (No significant difference for 1–2 months vs. < 1 month: EHR 1.02, 95% CI 0.93 to 1.11)

Colorectal				
Redaniel 2015[101]	Survival	T8	<>	
Breast				
Redaniel 2015[101]	Survival	T8	<>	
Lung				
Gildea 2017[118]	Stage (IIIb–IV)	T13	<> (NSCLC)	

Gonzalez-Barcala 2014[117]	Survival (3 years' follow-up)	T14	–	Longer intervals – better outcomes
				• The survival is longer in patients with a longer treatment delay
				• Interval in days
				  ○ ≥ 91 vs. 0–30 days: HR 0.64, 95% CI 0.43 to 0.96
				  ○ (No significant difference for –
				  – 31–60 vs. 0–30 days: HR 0.77, 95% CI 0.57 to 1.04
				  – 61–90 vs. 0–30 days: HR 0.66, 95% CI 0.42 to 1.02)

Gonzalez-Barcala 2014[117]	Survival (3 years' follow-up)	T15	–	Longer intervals – better outcomes
				• The survival is longer in patients with a longer treatment delay
				• Interval in days
				  ○ ≥ 61 vs. 0–30 days: HR 0.42, 95% CI 0.26 to 0.69
				  ○ (No significant difference for:
				  – 31–60 vs. 0–30 days: HR 0.81, 95% CI 0.54 to 1.11)

Gonzalez-Barcala 2014[117]	Survival (3 years' follow-up)	T3	<>	
Gonzalez-Barcala 2014[117]	Survival (3 years' follow-up)	T11	<>	

Kim 2016[119]	Stage (II, III vs. I)	T13	– (NSCLC)	Shorter intervals – worse outcomes
				• Outcome = delayed (≥ 38 days) 'diagnostic interval':
				  ○ Stage II vs. I: OR 0.68, 95% CI 0.50 to 0.92
				  ○ Stage III vs. I: OR 0.54, 95% CI 0.44 to 0.68

TABLE 53 Updated review: results of included studies[a] (*continued*)

Study	Outcome	Time interval[b]	Abbreviated results	Summary of results for studies reporting significant findings
Kim 2016[119]	Stage (III vs. I)	T15	– (NSCLC)	Shorter intervals – worse outcomes • Having locally advanced disease (stage III vs. I: OR 0.47, 95% CI 0.38 to 0.60) protected against delayed treatment intervals. (Stage II vs. I was not significant) • Outcome = delayed (≥ 51 days) 'treatment interval'
Kim 2016[119]	Stage (III vs. I)	T14	– (NSCLC)	Shorter intervals – worse outcomes • Factors associated with prompt care included stage III disease • Locally advanced stage (stage III vs. I: OR 0.42, 95% CI 0.33 to 0.53) was highly protective against delayed 'system intervals' and was not found to be collinear in the multivariable regression modelling. (Stage II vs. I was not significant) • Outcome = delayed (≥ 78 days) 'system interval'
Radzikowska 2013[115]	Survival	T1	<> (SCLC)	
Radzikowska 2013[115]	Survival	T8	– (SCLC)	Shorter intervals – worse outcomes • Patients who were diagnosed faster (< 42 days) had a worse prognosis than those diagnosed later. HR for death was 1.20 in group with interval of < 42 days (vs. > 42 days, $p = 0.001$)
Živković 2014[116]	Survival	T1	<>	
Živković 2014[116]	Survival	T4	<>	
Lymphoma				
Nikonova 2015[120]	Survival	T7	<>	
Nikonova 2015[120]	Survival	T14	<>	
Myeloma				
Goldschmidt 2016[121]	Survival	T8	<>	
Goldschmidt 2016[121]	Stage	T8	<>	
Neuroendocrine				
Keizer 2016[122]	Survival	T4	<>	
Keizer 2016[122]	Stage	T4	<>	

continued

TABLE 53 Updated review: results of included studies[a] (*continued*)

Study	Outcome	Time interval[b]	Abbreviated results	Summary of results for studies reporting significant findings
Oral				

Alahapperuma 2017[123]	Stage [III–IV (vs. I–II)]	T1	+	Longer intervals – worse outcomes
				• Stage at diagnosis was significantly associated with 'patient delay – 1' ($p = 0.001$) (delay defined as > 3 months)
				• Number (%) of patients with early or advanced stage for those with and without 'delay':
				  ○ Present: early, 6 (12.5%) vs. advanced, 42 (87.5%)
				  ○ Absent: early, 90 (44.1%) vs. advanced, 114 (55.9%)

Alahapperuma 2017[123]	Stage [III–IV (vs. I–II)]	T10	<>	
Esmaelbeigi 2014[124]	Stage (advanced)	T5	+	Longer intervals – worse outcomes • The delayed patients (interval > 140 days) were diagnosed in more advanced stage than the patients without delay (OR 2.1, 95% CI 1.0 to 4.4)
Ovarian				
Altman 2017[126]a	Survival	T8	+ (after 30 days)	• Longer intervals – no adverse outcomes • Time to diagnosis was significantly related to survival ($p = 0.0309$) in the analysis for late-stage disease, but was not identified as a significant factor for early-stage disease • OS was not negatively affected by longer time to diagnosis up until 80 days, at which point survival began decreasing with longer diagnostic interval • Emergency presentation was associated with a worse OS ($p = 0.0399$) in late-stage disease
Lim 2016[125]	Stage (advanced)	T1	<> (type I and II)	
Lim 2016[125]	Stage (advanced)	T8	<> (type I and II)	
Pancreatic				
Gobbi 2013[127]	Survival	T4	+	Longer intervals – worse outcomes Time to diagnosis (weeks) was a significant and independent prognostic factor: HR 1.02, 95% CI 1.01 to 1.04. There was a significant difference between the Kaplan–Meier curves for different thresholds ($p < 0.0001$): < 4, 4–6 and > 16 weeks
Jooste 2016[128]	Survival	T5	<>	

TABLE 53 Updated review: results of included studies[a] (*continued*)

Study	Outcome	Time interval[b]	Abbreviated results	Summary of results for studies reporting significant findings
Penile				
Gao 2016[129]	Survival	T1	+ (2-year OS) <> (5-year OS)	Longer intervals – worse outcomes • Patients with a delay of > 6 months showed significantly inferior 2-year OS • Interval – months o > 6 months vs. ≤ 1 month: OR 5.19, 95% CI 1.05 to 25.60 o (No significant difference for: – 1–3 months vs. ≤ 1 month: OR 0.98, 95% CI 0.13 to 7.21 – 3–6 months vs. ≤ 1 month: OR 2.9, 95% CI 0.51 to 16.59) • Patient's delay did not have an impact on 5-year OS
Gao 2016[129]	Stage (T1b–T4 vs. Tis/Ta/T1a)	T1	+	Longer intervals – worse outcomes • Patients with a delay of ≥ 3 months had significant risk for more advanced stage (T1b–T4) tumour (compared with a delay of ≤ 1 month) o 3–6 vs. months ≤ 1 month: OR 2.9, 95% CI 1.23 to 7.09 o > 6 months vs. ≤ 1 month: OR 6.42, 95% CI 2.99 to 13.79 o (No significant difference for 1–3 months vs. ≤ 1 month: OR 1.50, 95% CI 0.62 to 3.66)
Gao 2016[129]	Tumour size (> 2 cm)	T1	+	Longer intervals – worse outcomes • Patients with a delay of ≥ 3 months had significant risks for bigger (> 2 cm) lesion size than patients with a delay of ≤ 1 month o 3–6 months vs. ≤ 1 month: OR 3.29, 95% CI 1.53 to 7.07 o > 6 months vs. ≤ 1 month: OR 6.42, 95% CI 2.99 to 13.79 o (No significant difference for 1–3 months vs. ≤ 1 month: OR 1.43, 95% CI 0.67 to 3.02)
Gao 2016[129]	Metastases (M1)	T1	+	Longer intervals – worse outcomes • Patients with a delay of ≥ 3 months (vs. ≤ 1 month) had a significantly increased risk of metastasis
Gao 2016[129]	Lymph node involvement (N1–3)	T1	–	Longer intervals – worse outcomes Patients with a delay of > 6 months had significant risks for higher positive rate of regional lymph nodes (compared with patients with a delay of ≤ 1 month)

continued

TABLE 53 Updated review: results of included studies[a] (*continued*)

Study	Outcome	Time interval[b]	Abbreviated results	Summary of results for studies reporting significant findings
Prostate				
Bonfill 2015[130]	Stage (T2a–T4; reference T1a–T1c)	T8	<>	
Bonfill 2015[130]	Stage (T2a–T4; reference T1a–T1c)	T15	<>	
Sarcoma				
De Boer 2014[133]	Stage [poor-risk stage (vs. good risk)][c]	T1	+ (HIV Kaposi's sacoma)	Longer intervals – worse outcomes • Patients who experienced diagnostic delay (> 3 months) were more likely than those who did not delay (≤ 3 months) to have poor-risk Kaposi's sacoma stage (OR 3.41, 95% CI 1.46 to 7.45; $p = 0.002$)
Goedhart 2016[132]	Survival	T4	<>	
Goedhart 2016[132]	Survival	T11	<>	
Goedhart 2016[132]	Metastases	T4	<>	
Urakawa 2015[131]	Survival	T3	–	Shorter intervals – worse outcomes • Multivariate analysis revealed that symptom interval of < 6 months was an independent factor associated with poor survival ($p = 0.017$) ○ < 6 vs. ≥ 6 months: HR 3.93, 95% CI 1.29 to 11.97
Testicular				
Kobayashi 2014[134]	Survival	T1	+	Longer intervals – worse outcomes • Time to consultation of > 6 months was an independent risk factor associated with poorer OS (< 6 vs. ≥ 6 months, HR 18.0, 95% CI 1.78 to 182; $p = 0.014$)
Kobayashi 2014[134]	Stage (> II)	T1	<>	
Kobayashi 2014[134]	Size (larger; > 2 vs. ≤ 2 cm and > 5 vs. ≤ 5 cm)	T1	+	Longer intervals – worse outcomes • There was a close and positive correlation between time to consultation and primary tumour size (Kruskal–Wallis test; $p = 0.001$)

+, positive association, based on a statistically significant more favourable outcome; –, negative association, based on a statistically significant less favourable outcome; <>, findings that were not statistically significant; EHR, excess hazards ratio; HIV, human immunodeficiency virus; HR, hazard ratio; N, node; NSCLC, non-small-cell lung cancer; OR, odds ratio; OS, overall survival; SCLC, small-cell lung cancer.

a Ordered alphabetically by cancer site and then author, with studies evaluating more than one cancer reported under 'multiple' with each corresponding cancer site (the findings for these studies are presented in *Table 11* under the corresponding cancer site).

b See *Figure 2* for the time intervals.

c Using the validated staging system for AIDS-associated Kaposi sarcoma;[233] patients were classified as having overall "good risk" or "poor risk" based on this system. The diagnostic areas included were: extent of tumour involvement (T), immune system function (I), and presence of systemic illness (S). A patient with poor risk in all three areas, (T1I1S1), is defined as having overall poor risk. Good risk for tumour extent indicates that all KS nodules, or lesions, are confined to the skin or lymph nodes, and any oral involvement is confined to the palate only.

Colorectal cancer studies

TABLE 54 Updated review: categories used for coding study aim

Index relating to aim of study	Definition
Interval	Evaluate the influence of delay on patient outcomes (or association between interval and outcomes)
Delay	Describe the intervals (or identify delays). Includes comparison of intervals with recommended guidelines also
Progress	Changes in delay (intervals) over time
Symptoms	Evaluation of which specific symptoms are associated with delay
Causes	Identifying which factors are associated with delay (or determinants of delay). (Predictors of delay)
Prognostic factors	Identifying modifiable factors associated with patient outcomes (or identifying influence of various factors). (Predictors of stage/survival)
Pathway	Comparing different referral 'type' interventions or pathways, for example fast track vs. conventional; rapid/urgent referral vs. non-urgent
	Note that studies were excluded from the updated review for which this was the primary aim, and for which the reference details in the EndNote library did not indicate that the study also evaluated the impact of the interval on patient outcomes (survival or stage)

TABLE 55 Updated review: study characteristics for the included CRC studies

Author and year; country	Study aim category	Study design	Setting	Study population	Participants sampled (*n*)	Recruited or records available (*n*)	Analysed (*n*)	Data source	Interval	Outcome measure
Aslam 2017;[110] the UK	Pathway; interval	Cohort study (unclear if retrospective or prospective)	Secondary care (single site)	Patients diagnosed with CRC within the University Hospitals of Leicester NHS Trust, between 2005 and 2012. Interval comparison in patients undergoing surgery. Study also included patients referred via screening ($n = 275$), but not included in survival analysis. Included emergency admission ($n = 967$), but also presented survival analysis excluding these patients	3618 (47 missing)	3571 (967 emergency presentation; 275 screening)	2329	Database maintained by the NHS trust on CRC outcomes	T12	Survival
Chen 2017;[112] the USA	Delay; causes; interval	Retrospective comparative cohort	Secondary care (single site)	Patients diagnosed with colorectal adenocarcinoma, confirmed by pathologists, and seen at the Stanford Cancer Institute between 2008 and 2014. Compared data of patients with young-onset CRC (diagnosed at an age < 50 years) with those patients diagnosed with CRC at an age of ≥ 50 years. Excluded patients with insufficient data, and, for cohort ≥ 50 years, those who were asymptomatic. First medical visit was with a primary care provider for 70% (320/458) of patients, and emergency department for 25% (113/458)	1759 (382 diagnosed at age < 50 years; 1377 at age ≥ 50 years)	354 for those aged < 50 years (101 with insufficient data); 1276 for those aged ≥ 50 years, from which a random sample of 436 was taken (95 screening age, 109 with insufficient data)	455 (253 aged < 50 years; 232 aged ≥ 50 years)	Stanford Cancer Institute Research Database (clinical records) used to identify cases, and obtain data on interval and outcome	T4, T1, T8	Stage

Author and year; country	Study aim category	Study design	Setting	Study population	Participants sampled (*n*)	Recruited or records available (*n*)	Analysed (*n*)	Data source	Interval	Outcome measure
Dregan 2013;[108] the UK	Symptoms; interval	Retrospective population-based cohort study	Primary care (multicentre)	Patients presenting with a specific alarm symptom (haematuria, haemoptysis, dysphagia or rectal bleeding) or diagnosed with the corresponding cancer (urinary tract, lung, gastro-oesophageal, or colorectal, respectively) recorded between 2002 and 2006. Excluded patients with incomplete data or discrepant dates	5791 eligible (498 CRC)	5791 (267 excluded; 47 for CRC)	5524 (451 CRC)	Primary care records (UK CPRD) linked with Cancer Registry data	T4	Survival
Helewa 2013;[106] Canada	Interval; prognostic factors; delay	Retrospective population-based cohort study	Secondary care (multicentre)	Patients diagnosed with stage I–IV CRC, between 2004 and 2006, who underwent major surgical resections in Manitoba. Included emergency presentations ($n = 262$), which were also analysed separately. Patients for whom dates of index contact could not be determined were excluded, as were patients seen in emergency department	2086 diagnosed (1628 underwent surgery)	Diagnostic interval (and total wait time): diagnostic interval + treatment wait) could be determined for only 1307 out of 1628 (80.3%); treatment wait for 1628	Unclear (1628)	Population-based Manitoba Cancer Registry (linked with other administrative databases)	T9	Survival (adjusted for stage and emergency presentations)
Janssen 2016;[114] Canada	Delay; interval	Retrospective chart review	Secondary care (single site)	Patients referred to, and subsequently seen by, a gastroenterologist at St. Paul's Hospital, Vancouver, and had a pathological diagnosis of CRC, between 2010 and 2013. Patients with a mass on rectal/physical examination or a high suspicion of malignancy based on diagnostic imaging at the time of referral were excluded	592	246 met inclusion criteria, but node positivity and distant metastases could not be determined in 13 patients	Unclear (246)	Patients identified via hospital pathology database and then cross-referenced against EMR to select those seen by the gastroenterologists ($n = 10$) at the hospital	T11	Stage

continued

TABLE 55 Updated review: study characteristics for the included CRC studies (*continued*)

Author and year; country	Study aim category	Study design	Setting	Study population	Participants sampled (*n*)	Recruited or records available (*n*)	Analysed (*n*)	Data source	Interval	Outcome measure
Leiva 2017;[234] Spain	Delay; interval; causes	Retrospective population-based multicentre cross-sectional study	Primary and secondary care (multicentre)	Consecutive series of symptomatic patients diagnosed with CRC at one of nine public hospitals from 2006 to 2008 and registered with GP	795	Interval data available via: interviews for 715 (61 not interviewed, 19 no onset date), hospital records for 637, and GP records for 316	715 for 'patient recall', 637 for 'hospital records' and 316 for 'GP records'	Cases of CRC identified from pathology services. Data on intervals identified using three different sources of information: 1. Patient-recorded data by interview 2. hospital-recorded data 3. GP-recorded data	T4, T8	Stage
Murchie 2014;[107] the UK (Scotland)	Interval; prognostic factors	Retrospective population-based cohort study	Primary and secondary care (multicentre)	Symptomatic patients diagnosed with CRC between 1997 and 1998 in northern Scotland identified via CRUX data set (numbers not stated). Those with signs and symptoms > 2 years prior to treatment were excluded	NR	Unclear	958 for survival; 868 stage	Linked data from four data sets: CRUX data sets (primary care data), Scottish Cancer Registry (for tumour stage and grade), the Scottish Death Registry and the acute hospital discharge (SMR01) data set	T9	Survival and stage (adjusted for tumour grade, emergency admissions, and signs and symptoms)
Pita-Fernández 2016;[109] Spain	Interval; delay; symptoms	Retrospective cohort study	Tertiary care (single site)	Patients diagnosed with CRC at the Complexo Hospitalario Universitario A Coruña (A Coruña, north-west Spain) between 1994 and 2000	1482	Unclear; 942 with available data from clinical records to calculate interval	942	Interval data from clinical records with mortality data obtained from Galician Mortality	T4	Survival (adjusted for stage)
Patel, 2018;[111] the UK (Scotland)	Interval; delay	Retrospective cohort	Secondary care (multicentre)	Patients treated for CRC at one of three district general hospitals (1999–2005). Included patients presenting electively or in an emergency setting (with patients who did not go on to have surgery within 72 hours of admission reclassified as elective) ($n = 1012$)	1672 (660 missing data)	1012 (794 elective, 218 emergency)	1012	Patient records	T12	Survival; stage

Author and year; country	Study aim category	Study design	Setting	Study population	Participants sampled (*n*)	Recruited or records available (*n*)	Analysed (*n*)	Data source	Interval	Outcome measure
Pruitt 2013;[137] the USA	Interval; delay	Retrospective population-based cohort study (matched case–control)	Secondary care (multicentre)	Older US adults (aged ≥ 66 years) with invasive colon or rectal cancer and full Medicare coverage were eligible. Patients with a relevant Mediacaid claim and CRC diagnosed between 1998 and 2005 were included. Excluded patients with pre-existing comorbidities, and patients presenting for emergency procedures For T8, patientts without a claim for symptom presentation visit were excluded (presumed to be identified via screening), and for T15 those who did not have treatment within 12 days of diagnosis	Total NR; 9663 not eligible	10,663 eligible (excluded: 994 no symptom presentation claim; 979 no treatment)	9669 for T8 (6702 colon, 2967 rectal) and 9684 for T15 (6698 colon, 2986 rectal)	An existing linkage of the 1998–2005 National Cancer Institute's SEER programme data with 1997–2006 Medicare claims files from the Centers for Medicare and Medicaid Services	T8, T15	Survival (adjusted for stage, grade, location; stratified by symptom, stage)
Redaniel 2015;[101] the UK (England)	Interval	Retrospective population-based cohort study	Primary care (multicentre)	Patients diagnosed with colorectal, breast, lung or prostate cancers who presented to a GP with a cancer symptom 1 year prior to diagnosis identified. Only English cancer patients diagnosed in the years covered by all four data sets (1998–2009) were included. (Excluded patients diagnosed through emergency routes)	62,178 eligible patients (cancer diagnosis) on CPRD (45,766 linked to other databases)	22,051 (48%) patients presented to a GP with a cancer symptom 1 year prior to diagnosis (5912 with CRC)	22,051 (5912 CRC)	CPRD was linked to the NCDR and to the 2007 English IMD data sets. (NCDR captures data from the merged Cancer Registry, Hospital Episode Statistics and the ONS)	T8	Survival (stratified by alert or non-alert symptom; and adjusted for stage, tumour differentiation, and subsite, among other factors)

continued

TABLE 55 Updated review: study characteristics for the included CRC studies (*continued*)

Author and year; country	Study aim category	Study design	Setting	Study population	Participants sampled (*n*)	Recruited or records available (*n*)	Analysed (*n*)	Data source	Interval	Outcome measure
Tørring 2013;[99] Denmark	Interval; causes	Retrospective population-based cohort study	Primary care (multicentre)	Patients in the former Danish county of Aarhus with newly diagnosed colorectal, lung, melanoma skin, breast or prostate cancer during a 1-year period (2004–05), and whose GPs were involved in the diagnosis. (included emergency admission). (Note that the CRC cohort appears to be the same as is Tørring 2011[136])	1543	1376 involved GP diagnosis (GP did not participate for 248)	1128 (268 CRC)	Consecutive cancer patients identified from the County Hospital Discharge Registry, and linked to historical database hosted at the Department of Clinical Epidemiology, Aarhus University Hospital (via civil registry number). Each patient's GP was subsequently identified by linking the patient's data to the Health Service Registry. Same sources used for data on covariates, interval, and death Date of first presentation, symptoms, and date of diagnosis obtained via GP questionnaire (84% response rate)	T8	Survival (adjusted; stratified by alarm/vague symptom, cancer)
Tørring 2012;[135] the UK ($n = 1$) and Denmark ($n = 2$)	Interval	Three population-based cohort studies [one retrospective case–control studies (UK); and two prospective cohort studies]	Primary care (multicentre)	Analysed data from three previously described population-based studies in Denmark ($n = 2$) and the UK ($n = 1$). All newly diagnosed CRC patients aged > 39 years were included from each study. (Included emergency admission.) (Note that one of the studies, based on a GP questionnaire, is Tørring 2011[136])	1667 (361 for GP recall; 945 for 'patient recall'; 361 GP records)	• 1389 (missing data for 58 GP recall; 208 for patient recall; 12 GP records) • 1389 patients with data: 303 GP recall; 737 for patient recall; 349 GP records	1243 (266 GP-recall; 658 for patient recall; 319 GP records) Excludes 146 – GP not involved (37 GP recall; 79 for patient recall; 30 GP records)	• Cases identified via hospital discharge registry, hospital department, or local cancer registry (UK). (Separate validation source reported for two Danish studies) • Data on interval collected via primary care patient records (UK); patient interview questionnaire; and GP postal questionnaire • Survival data obtained from Civil registration data (Denmark) or local cancer registry (UK)	T8	Survival

Author and year; country	Study aim category	Study design	Setting	Study population	Participants sampled (*n*)	Recruited or records available (*n*)	Analysed (*n*)	Data source	Interval	Outcome measure
Tørring 2011;[136] Denmark	Interval	Prospective population-based cohort study	Primary care (multicentre)	All newly diagnosed CRC patients aged > 17 years during 1 year (1 September 2004 to 31 August 2005). Excluded patients whose GP was not involved in diagnosis. (Included emergency admission)	363	326 involved GP diagnosis (GP did not participate for 58)	268	• Cases identified via the Country Hospital Discharge Registry (diagnosis verified via Danish Cancer Registry), linked to County Health Service Registry to identify patients' GPs • Data on interval collected via GP questionnaire • Survival data obtained from registry data (Danish Civil Registration System)	T8	Survival

CRUX, Comparing Rural and Urban Cancer Care; EMR, emergency medical record; NCDR, National Cancer Data Repository; NR, not reported; ONS, Office for National Statistics.

Notes

Categories for study aim: interval – evaluate the influence of delay on patient outcomes; delay – describe the intervals (or identify delays); progress – changes in delay (intervals) over time; symptoms – evaluation of which specific symptoms are associated with delay; causes – identifying which factors are associated with delay; prognostic factors – identifying modifiable factors associated with patient outcomes; pathway – comparing different referral 'type' interventions or pathways.

BOX 1 Updated review: bias assessment tool

Sample representativeness

Is the sample representative of the relevant cancer patient population? The population may be quite specific, typified by age, stage, ethnicity or other factors.

Yes: only if this is clearly reported.

Can't tell: if it's reported in an ambiguous way.

Not reported: if it doesn't say.

If none of the above – please qualify with free text (this may be a majority of studies).

Characteristics' reporting

Was the reporting of participant characteristics complete?

Yes: only if this is clearly and fully reported.

Can't tell: is if it's reported in an ambiguous way.

Not reported: if it doesn't say.

If none of the above – please qualify with free text (this may be the majority of studies, e.g. those that give just age and sex).

Representativeness of participants

Were participants who participated (or whose data were used) representative of the sample from which they (or it) were sourced?

Yes: only if this is clearly and fully reported.

Can't tell: is if it's reported in an ambiguous way.

Not reported: if it doesn't say.

Not applicable: if all of sample participated – for example database study.

If none of the above – please qualify with free text.

Bias minimisation in measurement of symptom duration

Were steps taken (as stated by the investigators) to minimise and check for biases and inaccuracies introduced due to the method used for measurement of symptom duration?

Yes: if clear evidence of this, please list information as free text: MANDATORY.

No: if no evidence of this.

Can't tell: if unclear (this includes where results may be reported but no mention in methods).

BOX 1 Updated review: bias assessment tool (*continued*)

Independent variable assessment

Was the assessment symptom duration (explanatory variable) conducted independent of the assessment of the outcome variable?

Yes: if reported as done.

No: if clearly reported that same researcher did it.

Not reported: if it doesn't say.

Not applicable: if method does not require this to be done, for example database study.

A priori definition of outcome variable

Was the outcome variable specified/defined a priori?

Yes.

No.

Appropriate definition of outcome variable

Was the outcome variable clearly defined?

Yes: for example type of stage, type of survival – not necessary to enter detail.

No: anything other than yes.

Multivariate analysis

Was multivariate analysis conducted?

Yes.

No.

Prognostic adjustment

Was adjustment for important prognostic factors conducted as part of the analysis?

Yes: if clear evidence of this (e.g. performance status, age, smoking, comorbidity) please qualify with free text MANDATORY.

No: if no evidence of this.

Can't tell: if unclear (this includes studies for which results may be reported but no mention in methods) please qualify with free text MANDATORY.

BOX 1 Updated review: bias assessment tool (*continued*)

Outlier adjustment for symptom duration

Was adjustment for outliers conducted as part of the analysis?

Yes: if clear evidence of this – please qualify with free text MANDATORY.

No: if no evidence of this.

Not applicable: for example if there were no symptom durations greater than 2 years for more quickly diagnosed cancers.

Can't tell: if unclear (this includes studies for which results may be reported but no mention in methods) please qualify with free text MANDATORY.

Confounder adjustment

Was adjustment for confounders (identified in advance of the study) conducted as part of the analysis?

Yes: if clear evidence of this, please qualify with free text MANDATORY.

No: if no evidence of this.

Can't tell: if unclear (this includes where results may be reported but no mention in methods) please qualify with free text MANDATORY.

TABLE 56 Updated review: critical appraisal for the included CRC studies

Study	Time interval[a]	Sample representativeness	Characteristics reporting	Representativeness of participants	Bias minimisation	Independent variable assessment	A priori definition	Appropriate definition	Multivariate analysis	Prognostic adjustment	Outlier adjustment	Confounder adjustment
Aslam 2017[110]	T12	Patients receiving surgery	NR	Yes	No	N/A	Yes	Yes	No	No	No	No
Chen 2017[112]	T4, T1, T8	Yes	Yes	Cannot tell	No	N/A	Yes	Yes	No (for outcome of interest)	No	No	No
Dregan 2013[108]	T4	Yes	Just age and sex	Yes	Used sensitivity analysis[b]	N/A	Yes	Yes	Yes	No	No	No
Helewa 2013[106]	T9	Patients receiving surgery	Yes	Cannot tell	No	N/A	Yes	Yes	Yes	Yes	No	Yes
Janssen 2016[114]	T11	Yes	Yes	Yes	Yes	N/A	Yes	Yes	No	No	N/A	No
Leiva 2017[113]	T8, T4	Yes	Yes	Cannot tell	No	NR	Yes	Yes	Yes	Yes	No	Yes
Murchie 2014[107]	T9	Yes	Yes	Cannot tell	No	N/A	Yes	Yes	Yes	Yes	N/A	Yes
Patel 2018[111]	T12	Yes	Just age and ASA grade[c]	Cannot tell	No	N/A	Yes	Yes	No	No	No	No
Pita-Fernández 2016[109]	T4	Yes	Yes	Cannot tell	No	NR	Yes	Yes	Yes	No (age, sex and stage only)	No	Yes
Pruitt 2013[137]	T8, T15	Older Medicaid patients	Yes	Yes	Yes[d]	N/A	Yes	Yes	Yes	Yes (histology, grade, stage, location)	N/A	Yes
Redaniel 2015[101]	T8	Yes	Yes	Yes	Cannot tell	N/A	Yes	Yes	Yes	Yes	N/A	Yes
Tørring 2013[101]	T8	Yes	Yes	Yes (16% excluded)	No	N/A	Yes	Yes	Yes	Yes (age, sex and comorbidity only)	No	Yes
Tørring 2012[135]	T8	Yes	Yes	Yes	Yes	N/A	Yes	Yes	Yes	No	Unclear	Yes
Tørring 2011[136]	T8	Yes	Yes	Yes	Unclear	N/A	Yes	Yes	Yes	No	Unclear	Yes

ASA, American Society of Anesthesiologists; GPRD, General Practice Research Database; N/A, not applicable; NR, not reported.

a See *Figure 2* for the time intervals.

b Date of first alarm symptom obtained from database of GP records (GPRD) and date of diagnosis (and death) from the Cancer Registry or CPRD. There were some discrepancies between date of diagnosis in the Cancer Registry and the GPRD. Sensitivity analysis conducted ignoring cases for whom data of diagnosis was prior to data of first alarm symptom (did not alter results).

c ASA grade is a widely used subjective assessment of comorbidity employed largely by anaesthetists to label anaesthetic risk.

d To test for any bias resulting from the selected cut-off points and reference categories, all models were reanalysed using different categorisations of delay (including model-specific quartiles and quintiles) and different reference categories.

TABLE 57 Updated review: results of CRC studies

Study	Outcome	Time interval[a]	Association	Outcome description	Statistic test/method of analysis	Results
Aslam 2017[110]	Survival	T12	<>	Median OS	Log-rank test comparing median survival from Kaplan–Meier curves. The study also included patients referred via screening, survival analysis limited to emergency, 2WW, urgent, and routine referrals. Analysis also conducted excluding emergency	There was no significant difference in OS between the two interval groups (for 2WW, urgent, and routine referrals combined): < 62 days group: 7.1 years vs. > 62 days group: 6.54 years ($p = 0.620$)

Chen 2017[112]	Stage	T4, T1, T8	– (Aged < 50 years)
			– (Aged < 50 years)
			<>
				TNM stage: advanced (stage III–IV) vs. non-advanced (stage I–II)	The median interval (days) from symptom onset and/or workup to diagnosis was compared for advanced and non-advanced stage at diagnosis (but not statistically) within each age group	The authors noted that the longest time to diagnosis (T4) was observed in patients aged < 50 years with non-advanced CRC at diagnosis (stage I and II) followed by patients aged < 50 years with advanced CRC at diagnose (stages III and IV) (median days 174 vs. 124)
						• Aged < 50 years, non-advanced vs. advanced stage
						  ○ Median symptom duration (T1): 90 vs. 60 days
						  ○ Median workup duration (T8): 39 vs. 29 days
						• Aged ≥ 50 years, non-advanced vs. advanced stage:
						  ○ Median symptom duration (T1): 21 vs. 30 days
						  ○ Median workup duration (T8): 31 vs. 17 days

Study	Outcome	Time interval[a]	Association	Outcome description	Statistic test/method of analysis	Results
Dregan 2013[108]	Survival	T4	<>	Overall mortality (from diagnosis)	Adjusted HR (95% CI) obtained from Cox regression model (adjusted for age and sex). (Risk of death according to presence or absence of alarm symptoms was also analysed)	• ≤ 14 days ($n = 79$), 1.07 (95% CI 0.65 to 1.77); $p = 0.794$ • 15–90 days ($n = 217$), reference category • 91–180 days ($n = 56$), 1.28 (95% CI 0.73 to 2.25); $p = 0.386$ • 181–365 days ($n = 44$), 0.76 (95% CI 0.38 to 1.54); $p = 0.448$ • > 365 days ($n = 55$), 0.92 (95% CI 0.45 to 1.85); $p = 0.806$ (Patients with no preceding alarm symptoms had shorter survival from diagnosis than those presenting with relevant alarm symptoms)
Helewa 2013[106]	Survival	T9	<>	5-year OS	Adjusted HR (95% CI) and p-values obtained from multivariate Cox proportional hazards model using bootstrap resampling. Important confounders that were adjusted for included, among others, stage, emergency presentations and location. Emergency presentations were also analysed separately. An analysis limited to stage I–III CRC was also conducted (excluding stage IV and 'unknown')	Increased total wait time quartile was not associated with worse survival ($p = 0.4898$) • Q1: 0 to ≤ 43 days ($n = 175$), reference • Q2: > 43 to ≤ 95 days ($n = 135$): 0.93 (95% CI 0.73 to 1.19) • Q3: > 95 to ≤ 166 days ($n = 115$): 0.90 (95% CI 0.69 to 1.16) • Q4: > 166 to ≤ 513 days ($n = 113$): 0.82 (95% CI 0.64 to 1.06) When total wait time was considered as a continuous variable (days), it was still not associated with worse survival ($p = 0.2618$)
Janssen 2016[114]	Stage (advanced)	T11	<>	Node positivity, and presence of distant metastases	Chi-squared tests	Node positive in < 60 days group: 42/102 (41%) vs. > 60 days group: 52/144 (36%) ($p = 0.42$) Distant metastases in < 60 days group: 12/102 (12%) vs. > 60 days group: 9/144 (6%) ($p = 0.13$)

continued

TABLE 57 Updated review: results of CRC studies (*continued*)

Study	Outcome	Time interval[a]	Association	Outcome description	Statistic test/method of analysis	Results
Leiva 2017[113]	Stage (advanced)	T8	– (Hospital) <> (GP)	TNM tumour stage (0, 1, 2, 3, 4, unknown)	The Mann–Whitney *U*-test and the Kruskal–Wallis test were used to assess the association between diagnostic intervals and various patient and clinical characterises (including, among others, stage, emergency presentation, location, symptoms at presentation and perceived seriousness of symptoms). The independent effect of variables was assessed using Cox regression and then the extension of the multivariate Cox proportional hazards model with time-dependent covariates	• Comparing three different information sources for estimating 'diagnostic intervals': patient recall, hospital records and GP records. Interval defined as date from first registry of symptoms to date of diagnosis for GP and hospital records (T8), but for patient recall this was time from patient first experiencing symptoms to date of date of diagnosis (T4) • Univariate analysis indicated that a shorter diagnostic interval was associated with more advanced tumour stage using both GP and hospital records, but this was only significant for hospital records' data ($p = 0.021$) • Stage was not an independent predictor of a shorter interval in the multivariate analysis of hospital and GP data
		T4	<> (patient)	TNM tumour stage (0, 1, 2, 3, 4, unknown)	As above	• Univariate analysis indicated that a shorter diagnostic interval was associated with more advanced tumour stage for patient recall, but this was not significant • Stage was identified as one of the independent predictors of shorter interval in the multivariate analysis for patient-recorded data (but not in the analysis of hospital and GP data)

Study	Outcome	Time interval[a]	Association	Outcome description	Statistic test/method of analysis	Results
Murchie 2014[107]	Stage (advanced)	T9	<>	Dukes' stage was collapsed into a binary variable: early (A or B) vs. advanced (C or D)	OR (99% CI) obtained from logistic regression model with a restricted cubic spline (to allow for non-linear relationship), following sequential adjustment (using 1–4 models; 1 representing unadjusted) of patient and tumour factors (including, among others, tumour grade, and signs and symptoms). Time to treatment was modelled using four knots corresponding to the 25th, 50th, 75th and 100th centile	The spline curves showed a significant non-linear ($p = 0.04$) (and unadjusted, $p = 0.002$) association between provider delay and stage. Delays of between 4 and 34 weeks were associated with earlier-stage disease, and intervals beyond this were associated with later-stage disease. However, the plot based on a fully adjusted model showed wide CIs According to the fully adjusted model, provider delays of 40 and 60 weeks were associated with later-stage disease at presentation, but the OR did not reach statistical significance • 4 weeks: reference • 20 weeks: 0.87 (95% CI 0.54 to 1.39) • 40 weeks: 1.46 (95% CI 0.93 to 2.31) • 60 weeks: 1.58 (95% CI 0.96 to 2.61)
Murchie 2014[107]	Survival	T9	<>	All-cause survival (from date of first presentation)	HR (99% CI) obtained from Cox survival models, both with a restricted cubic spline (to allow for non-linear relationship), following sequential adjustment (1–4 models) of patient and tumour factors (including, among others, tumour grade, emergency admission, and signs and symptoms). Time to treatment was modelled using four knots corresponding to the 25th, 50th, 75th and 100th centile. Stratified analysis also conducted for rectal and colon cancer	The spline curves showed a significant non-linear ($p < 0.001$) (and unadjusted inverse, $p < 0.001$) association between provider delay and mortality. In the univariate analysis, diagnostic interval of < 4 weeks was associated with poor survival, but this was no longer present after adjusting for confounders. According to the fully adjusted model, provider delays of 40 and 60 weeks were associated with later-stage disease at presentation, but the OR did not reach statistical significance • 4 weeks: reference • 20 weeks: 0.99 (95% CI 0.76 to 1.27) • 40 weeks: 1.17 (95% CI 0.92 to 1.48) • 60 weeks: 1.21 (95% CI 0.94 to 1.57)

continued

TABLE 57 Updated review: results of CRC studies (*continued*)

Study	Outcome	Time interval[a]	Association	Outcome description	Statistic test/method of analysis	Results
Patel 2018[111]	Survival	T12	–	Long-term (mean) survival	Proportion of patients still alive at long-term follow-up (October 2011), and log-rank test comparing Kaplan–Meier curves; 30-day postoperative mortality and cause of death were also assessed	Long-term survival was greatest in elective patients who did not receive treatment > 62 days (52% alive), compared with elective cases treated ≤ 62 days (34% alive). This was supported by the log-rank analysis ($p < 0.001$) Operative mortality was higher in patients treated ≤ 62 days (7% elective, 20% emergency) than in those who were treated > 62 days (4% elective, 7% emergency). The most common cause of death was CRC in all groups
Patel 2018[111]	Stage	T12	–	Stage (Dukes' stage A–B vs. C–D)	Proportions compared using chi-squared test	Proportion of early-stage disease (Dukes' stage A–B) was highest in elective patients treated > 62 days (50%) ($p < 0.01$) and lowest in emergencies treated ≤ 62 days (30%) ($p = 0.26$) Later-stage disease (Dukes' stage C–D) was most common in emergency patients treated ≤ 62 days (58%) and lowest in elective patients treated > 62 days (36%)

Study	Outcome	Time interval[a]	Association	Outcome description	Statistic test/method of analysis	Results
Pita-Fernández 2016[109]	Survival	T4	<> (Rectal, – not adjusted for stage) <> (Colon) <> CRC	5-year mortality (after diagnosis)	The influence survival was analysed in two ways: 1. Kaplan–Meier survival curves were computed for each interval quartile, and compared using the log-rank test 2. The interval was treated as a continuous variable using restricted cubic splines with four knots and using the 50th percentile (3.4 months) as reference point The 5-year HRs (95% CIs) were estimated as a function of the duration of the delay interval and adjusted for age and sex (model 1), and subsequently stage (model 2), using proportional hazard Cox regression	• The Cox regression model adjusting for age and sex showed that, in rectum cancers, patients within the first interval quartile had lower survival ($p = 0.003$), but this was no longer statically significant when also adjusting for stage ($p = 0.084$) • No significant differences were found for colon cancer when adjusting for age and sex ($p = 0.282$) or also adjusting for stage ($p = 0.160$) • Longer intervals were not associated with poor survival in CRC patients. From age-, sex- and stage-adjusted model: o < 1.5 months: reference o 1.5–3.4 months: 0.97 (95% CI 0.71 to 1.32) o 3.4–6.4 months: 0.76 (95% CI 0.55 to 1.04) o > 6.4 months: 0.77 (95% CI 0.56 to 1.05) • The cubic splines regression analysis revealed that, for rectum cancer, 5-year mortality progressively increases for intervals less than the median (3 months) and decreases as the delay increases until approximately 8 months. In colon cancer, no significant relationship was found between interval and survival
Pita-Fernández 2016[109]	Stage (advanced)	T4	<>	TNM tumour stage (I, II–III, IV)	The association between the interval (categorised into quartiles) and tumour stage analysed using multivariate logistic regression (adjusting for age and sex)	The interval was not found to be significantly associated with stage at diagnosis (stage I: 3.6 months, stage II–III: 3.4 months, stage IV: 3.2 months; $p = 0.728$), even after adjusting for age and sex

continued

TABLE 57 Updated review: results of CRC studies (*continued*)

Study	Outcome	Time interval[a]	Association	Outcome description	Statistic test/method of analysis	Results
Pruitt 2013[137]	Survival	T8	+ (Colon) <> (Rectal)	All-cause mortality, and CRC-specific mortality	Cases (CRC deaths) and controls (deaths due to other causes or censored) were matched on survival time, and the association of delay with death was examined using logistic regression. ORs (95% CI) were estimated using multivariate regression analysis adjusting for sociodemographic, tumour (histology, stage, grade, location) and treatment factors. Stratified analysis also conducted according to and stage (and presenting symptom in sensitivity analysis)	• Colon cancer patients with the longest diagnostic delays (8–12 months vs. 14–59 days) had higher odds of all-cause (OR 1.31, 1.08 to 1.58), but not CRC-specific, death • Among rectal cancer patients, delays were not associated with risk of all-cause or CRC-specific death
Pruitt 2013[137]	Survival	T15	– (Colon, < 1 week) <> (Rectal)	All-cause mortality, and CRC-specific mortality	OR (95% CI) estimated using multivariate regression analysis adjusting for sociodemographic, tumour (histology, stage, grade, location), and treatment factors. Stratified analysis also conducted according to stage (and presenting symptom in sensitivity analysis)	• Colon cancer patients with the shortest treatment delays (< 1 week vs. 1–2 weeks) had higher odds of all-cause (OR 1.23, 95% CI 1.01 to 1.49), but not CRC-specific, death • Among rectal cancer patients, delays were not associated with risk of all-cause or CRC-specific death
Redaniel 2015[101]	Survival	T8	<>	5-year survival	5-year EHRs (95% CIs) were computed using a generalised linear model with a Poisson error structure. Univariable and multivariable models built, for each cancer site, and stratified by the nature of the symptoms (NICE-qualifying alert vs. non-alert). Multivariable models controlled for, among others, the effects of period of cancer plan implementation, Dukes' stage, tumour subsite (for CRC) and tumour differentiation	There was no evidence of an association between diagnostic interval and mortality. EHR based on multivariate analysis for all patients: • < 1 month: reference • 1–2 months: 0.94 (95% CI 0.82 to 1.07) • 3–6 months: 0.92 (95% CI 0.82 to 1.04) • > 6 months: 0.93 (95% CI 0.81 to 1.07) The results showed significant lower mortality associated with longer diagnostic intervals for patients presenting with non-alert symptoms (EHR > 6 months vs. < 1 month: 0.85, 95% CI 0.72 to 1.00; $p = 0.049$), whereas no significant associations were identified for patients presenting with NICE-qualifying alert symptoms

Study	Outcome	Time interval[a]	Association	Outcome description	Statistic test/method of analysis	Results
Tørring 2013[99]	Survival	T8	– (Alarm symptoms, Q1)	5-year overall mortality	• Analyses stratified according to the GP's interpretation of the presenting symptoms: alarm or serious vs. vague. OR (95% CI) obtained from logistic regression using restricted cubic splines and adjusting for comorbidity, age and sex. Each model tested against a model with no diagnostic interval term using the Wald test • Data on tumour stage and emergency admission collected, and used to describe patient characteristics and compare the two groups, with alarm/serious symptoms or vague symptoms • (Note that this appears to be the same CRC cohort as Tørring 2011136)	Alarm/serious symptoms ($n = 201$): the categorical analysis showed that both very short and long diagnostic intervals (first or fourth quartile compared with second and third) were associated with poor survival, but only the adjusted OR for the former was statistically significant – • Q1: 4.74 (95% CI 2.20 to 10.19) • Q2 + Q3: reference • Q4: 2.01 (95% CI 0.93 to 4.36) This association was confirmed by the spline regression analyses, which revealed convex (U-shaped) associations: The risk of dying within 5 years decreased with longer diagnostic intervals up to approximately the 60th percentile, and then increased ($p = 0.001$)
			<> (Vague symptoms)			Vague symptoms ($n = 67$): the categorical analysis showed that the first diagnostic interval quartile was associated with poor survival, compared with second and third, whereas the fourth was not, but these findings were not statistically significant. Adjusted ORs (95% CIs) for each quartile: • Q1: 0.74 (95% CI 0.19 to 2.80) • Q2 + Q3: reference • Q4: 1.08 (95% CI 0.28 to 4.12) The corresponding spline regression analyses did not show a significant association ($p = 0.620$)

continued

TABLE 57 Updated review: results of CRC studies (*continued*)

Study	Outcome	Time interval[a]	Association	Outcome description	Statistic test/method of analysis	Results
Tørring 2012[135]	Survival	T8	−/+ (GP based; Q1, Q4)	5-year overall mortality rate (after diagnosis)	Data taken from three studies using different methods for identifying date of first presentation. Data analysed for each study separately and combined The data were modelled in two ways: 1. Comparison of the first and fourth diagnostic interval quartiles with the second + third 2. Diagnostic interval rescaled and treated as a continuous variable using restricted cubic splines with four knots, with the 50th percentile used as the reference point The 5-year HRs (95% CIs) estimated as a function of the length of the diagnostic interval and adjusted for tumour site (colon/rectum), age and sex using proportional hazard Cox regression. In the analysis of the combined data, differences were also allowed for in study-specific baseline hazards. (Note that the GP record-based study is Tørring 2011[136])	GP-based study ($n = 266$) adjusted HRs for each quartile: • Q1: 1.73 (95% CI 1.16 to 2.59) • Q2 + Q3: reference • Q4: 1.75 (95% CI 1.17 to 2.63) Patient-based study ($n = 658$) adjusted HRs for each quartile: • Q1: 1.26 (95% CI 0.95 to 1.66) • Q2 + Q3: reference • Q4: 1.11 (95% CI 0.85 to 1.46) Record-based study ($n = 319$) adjusted HRs for each quartile: • Q1: 1.20 (95% CI 0.84 to 1.71) • Q2 + Q3: reference • Q4: 1.30 (95% CI 0.91 to 1.86) Combined data ($n = 1243$) adjusted HRs for each quartile: • Q1: 1.33 (95% CI 1.10 to 1.61) • Q2 + Q3: reference • Q4: 1.28 (95% CI 1.06 to 1.55) The association between diagnostic interval and survival was the same for all three types of data: displaying a U-shaped association with decreasing and subsequently increasing mortality with longer diagnostic intervals (i.e. U-shaped association observed using three different data collection methods in different health-care systems and over different time periods)
			<> (Patient based; GP records-based)			
			−/+ (combined; Q1, Q2)			

Study	Outcome	Time interval[a]	Association	Outcome description	Statistic test/method of analysis	Results
Tørring 2011[136]	Survival	T8	– (Alarm symptoms)	3-year overall mortality (after diagnosis)	• Analyses stratified according to the GP's interpretation of the presenting symptoms: alarm or serious vs. vague. Logistic regression used to estimate 3-year mortality ORs (95% CIs) as a function of the diagnostic interval using restricted cubic splines and adjusting for tumour site, comorbidity, age and sex. Also reported outcome data for 1-year mortality • Data on tumour stage and emergency admission collected, and used to describe patient characteristics and compare the two groups, with alarm/serious symptoms or vague symptoms	Alarm/serious symptoms ($n = 201$): The cubic spline regression analysis revealed that the risk of dying within 3 years decreased with diagnostic intervals of up to 5 weeks and then increased ($p = 0.002$). Adjusted OR for diagnostic intervals: • 0–4 weeks ($n = 75$): 2.56 (95% CI 1.29 to 5.05) • 5–11 weeks ($n = 90$): reference • ≥ 12 weeks ($n = 36$): 2.04 (95% CI 0.87 to 4.77)
			<> (Vague symptoms)			Vague symptoms ($n = 67$): the cubic spline regression analysis revealed a reverse effect, with increasing risk of dying from day 1 up to ≈ 12 weeks, but the association was not statistically significant ($p = 0.205$). Adjusted OR for diagnostic intervals: • 0–4 weeks ($n = 10$): comparison not justified • 5–11 weeks ($n = 27$): reference • ≥ 12 weeks ($n = 30$): 0.71 (95% CI 0.32 to 2.91)

+, positive association, whereby shorter intervals are associated with statistically significant more favourable outcomes; –, negative association, whereby shorter intervals are associated with statistically significant unfavourable outcomes; <>, no association, meaning the findings were not statistically significant; EHR, excess hazard ratio; HR, hazard ratio; OR, odds ratio; OS, overall survival; Q, quartile.

a See *Figure 2* for the time intervals.

Note

Favourable outcomes = improved survival and earlier stage at diagnosis.

TABLE 58 Summary results from previous review by Neal *et al.*[25]

Interval and outcome measure	Time interval[a]	Positive association	No association	Negative association
Breast				
Survival				
Diagnostic interval	T8	Tørring 2011[136]		
Treatment interval	T15	• Smith 2013[235] • Yun 2012[236]	• Brazda 2010[237] • Eastman 2013[238] • McLaughlin 2012[239] • Mujar 2013[240] • Redaniel 2013[241] • Sue 2013[242]	
Stage				
Symptom onset to diagnosis	T4	• Ermiah 2012[243] • Warner 2012[244]		
Treatment interval	T15		• Wright 2010[245] • Wagner 2011[246]	
Other				
Treatment interval and risk of recurrence	T15		Eastman 2013[238]	
Bladder				
Survival				
Symptom onset to referral	T2	Wallace 2002[247]		
Symptom onset to diagnosis	T4	Hollenbeck 2010[248]		
Symptom onset to treatment	T5		• Mommsen 1983[249] • Wallace 2002[247]	
From referral to first seen by specialist	T10		Gulliford 1991[250]	
From referral to treatment	T12		Gulliford 1991[250]	
From specialist care to treatment	T14		Gulliford 1991[250]	Wallace 2002[247]
Stage				
Patient interval	T1		Tokuda 2009[251]	
Symptom onset to diagnosis	T4		Maguire 1994[252]	
Diagnostic interval	T8	Liedberg 2003[253]		
Colorectal				
Survival				
Symptom onset to treatment	T5		• Thompson 2011[254] • Terhaar sive Droste 2010[255] – early-stage CRC	Terhaar sive Droste 2010[255] – late-stage CRC
Referral interval	T6		• Neal 2007[256] • Currie 2012[257] • Zafar 2012[258]	

TABLE 58 Summary results from previous review by Neal *et al.*[25] (*continued*)

Interval and outcome measure	Time interval[a]	Positive association	No association	Negative association
Diagnostic interval	T8	• Tørring 2011[136] • Tørring 2012[135] • Tørring 2013[99] • Pruitt 2013[137] – colon	Pruitt 2013[137] – rectal	
Treatment interval	T15	• Gort 2010[259] – colon • Yun 2012[236] – rectal	• Pruitt 2013[137] – rectal • Roland 2013[260]	Pruitt 2013[137] – colon
Stage				
Patient interval	T1		Van Hout 2011[261]	
Symptom onset to referral	T2		Van Hout 2011[261]	
Symptom onset to first seen by specialist	T3		Van Hout 2011[261]	
Symptom onset to diagnosis	T4		• Van Hout 2011[261] • Deng 2012[262]	
Symptom onset to treatment	T5		• Van Hout 2011[261] • Teehar sive Droste 2010[255] • Cerdán-Santacruz 2011[263]	
Referral interval	T6	• Valentín-López 2012[264] • Ramsay 2012[265] (routine referral group) • Guzman 2011[266] – colon	Neal 2007[256]	• Ramsay 2012[265] • Guzman 2011[266] – rectal
Treatment interval	T15	Guzman 2011[266] – colon		Guzman 2011[266] – rectal
Other				
Patient interval and satisfaction	T1		Tomlinson 2012[267]	
Lung				
Survival				
Patient interval	T1		Brocken 2012[268]	Radzikowska 2012[269]
Symptom onset to seen in specialist care	T3		Loh 2006[270]	
Symptom onset to diagnosis	T4	Maguire 1994[252]		
Symptom onset to treatment	T5		Annakkaya 2007[271]	
Referral interval	T6		Brocken 2012[268]	Neal 2007[256]
Diagnostic interval	T8	Tørring 2013[99]	• Skaug 2011[272] • Pita-Fernández 2003[273]	
From referral to first seen by specialist	T10		Brocken 2012[268]	
Seen in specialist care to diagnosis	T13		Brocken 2012[268]	Gould 2008[274]

continued

TABLE 58 Summary results from previous review by Neal *et al.*[25] (*continued*)

Interval and outcome measure	Time interval[a]	Positive association	No association	Negative association
Seen in specialist care to treatment	T14		Loh 2006[270]	Gould 2008[274]
Treatment interval	T15		• Brocken 2012[268] • Diaconescu 2011[275] • Yun 2012[236]	Gonzalez-Barcala 2010[276]
Stage				
Patient interval	T1		• Brocken 2012[268] • Tokuda 2009[251] • Yilmaz 2008[277]	
Symptom onset to treatment	T5	Christensen 1997[278]	Yilmaz 2008[277]	Myrdal 2004[279]
Referral interval	T6		Brocken 2012[268]	Neal 2007[256]
First seen in primary care to specialist	T7		Yilmaz 2008[277]	
Diagnostic interval	T8		Pita-Fernández 2003[273]	
Seen in primary care to treatment	T9		Yilmaz 2008[277]	
From referral to first seen by specialist	T10		Brocken 2012[268]	
Seen in specialist care to diagnosis	T13		Brocken 2012[268] Yilmaz 2008[277]	Gould 2008[274]
Seen in specialist care to treatment	T14			Gould 2008[274]
Treatment interval	T15	Murai 2012[280]	Brocken 2012[268] Yilmaz 2008[277]	Salomaa 2005[281]
Other				
Symptom onset to diagnosis, and quality of life	T4		Mohan 2006[282]	
Leukaemia *Survival*				
Symptom onset to diagnosis	T4		Prabhu 1986[283] – chronic myeloid	
Diagnostic interval	T8		Friese 2011[284] – chronic lymphocytic	
Treatment interval	T15		Bertoli 2013[285] – acute myeloid	
Lymphoma *Survival*				
Symptom onset to diagnosis	T4		• Jacobi 2008[286] – follicular • Maguire 1994[252] • Norum 1995[287] – Hodgkin's	Foulc 2003[288] – Sézary syndrome Kim 1995[289] – Sézary syndrome

TABLE 58 Summary results from previous review by Neal *et al.*[25] (*continued*)

Interval and outcome measure	Time interval[a]	Positive association	No association	Negative association
Myeloma				
Survival				
Symptom onset to diagnosis	T4	Kariyawasan 2007[290]		
Other				
Symptom onset to diagnosis and complications at diagnosis	T4	• Friese 2009[291] • Kariyawasan 2007[290]		
Neuroendocrine				
Stage				
Patient interval	T1		Tokuda 2009[251]	
Oral/head/neck				
Survival				
Patient interval	T1	• Koivenun 2001[292] – pharyngeal • Teppo 2008[293] – pharyngeal and laryngeal separately	• Teppo 2003[294] – laryngeal • Teppo 2008[293] – tongue	
Symptom onset to diagnosis	T4		• Wildt 1995[295] – oral • Seoane 2010[296] – oral	
Symptom onset to treatment	T5	Hansen 2005[297] – nasophayngeal	McGurk 2005[298] – head and neck	
Diagnostic interval	T8	• Alho 2006[299] – head and neck • Teppo 2003[294] – laryngeal • Teppo 2008[293] – laryngeal	• Teppo 2008[293] – pharyngeal and tongue • Koivunen 2001[292] – pharyngeal	
Treatment interval	T15	Sidler 2010[300] – nasopharyngeal	• Caudell 2011[301] – head and neck • Brouha 2000[302] – laryngeal	
Stage				
Patient interval	T1	• Kumar 2001[303] – oral • Brouha 2005[304] – pharyngeal and laryngeal separately • Sheng 2008[305] – nasopharyngeal • Tromp 2005[306] – unspecified • Tokuda 2009[251] – unspecified	• Allison 1998[307] – aerodigestive tract • Al-Rajhi 2009[308] – nasopharyngeal • Brouha 2005[304] – laryngeal • Wildt 1995[295] – oral • Teppo 2009[309] – vestibular schwannoma	
Symptom onset to referral	T2	Pitchers 2006[310] – oropharyngeal	Vernham 1994[311] – unspecified	
Symptom onset to seen in specialist care	T3	Allison 1998[307] – aerodigestive tract		

continued

TABLE 58 Summary results from previous review by Neal *et al.*[25] (*continued*)

Interval and outcome measure	Time interval[a]	Positive association	No association	Negative association
Symptom onset to diagnosis	T4	• Al-Rajhi 2009[308] – nasopharyngeal • Lee 1997[312] – nasopharyngeal	• Miziara 1998[313] – laryngeal • Scott 2005[314] – oral	
Symptom onset to treatment	T5		McGurk 2005[298] – unspecified	
First seen in primary care to specialist	T7	Allison 1998[307] – aerodigestive tract		
Diagnostic interval	T8	Al-Rajhi 2009[308] – nasopharyngeal	• Teppo 2009[309] – vestibular schwannoma • Ho 2004[315] – oropharyngeal	
Other				
Patient interval and risk of recurrence	T1		Teppo 2005[316] – laryngeal	
Diagnostic interval and risk of recurrence	T8	Teppo 2005[316] – laryngeal		
Cervical				
Survival				
First seen in specialist care to treatment	T14		Umezu 2012[317]	
Stage				
Patent interval	T1	Fruchter 1981[318]	Tokuda 2009[251]	
Symptom onset to diagnosis	T4	Fruchter 1981[318]		
Diagnostic interval	T8		Fruchter 1981[318]	
Endometrial				
Survival				
Symptom onset to diagnosis	T4		Menczer 1995[319]	
Referral to treatment	T12			Crawford 2002[320]
Treatment interval	T15			Elit 2003[321]
Stage				
Patent interval	T1		Tokuda 2009[251]	
Symptom onset to diagnosis	T4	• Franceschi 1983[322] • Fruchter 1981[318] • Obermair 1996[323]	Pirog 1997[324]	
Other				
Symptom onset to treatment, quality of life and satisfaction	T5	Robinson 2012[325]		

TABLE 58 Summary results from previous review by Neal *et al.*[25] (*continued*)

Interval and outcome measure	Time interval[a]	Positive association	No association	Negative association
Ovarian				
Survival				
Symptom onset to diagnosis	T4		Nagle 2011[326]	
Referral interval	T6		Neal 2007[256]	
Stage				
Patient interval	T1		• Smith 1985[327] • Tokuda 2009[251]	
Symptom onset to diagnosis	T4		• Fruchter 1981[318] • Menczer 2009[319] • Nagle 2011[326]	Lurie 2012[328]
Other				
Symptom onset to treatment and QOL and satisfaction	T5	Robinson 2012[325]		
Pancreatic				
Survival				
Symptom onset to referral	T2	Raptis 2010[329]		
Symptom onset to diagnosis	T4	Gobbi 2013[127]		
Treatment interval	T15		Yun 2012[236]	
Stage				
Patient interval	T1		Tokuda 2009[251]	
Other				
Diagnostic interval and resectability	T8		McLean 2013[330]	
Upper tract urothelial				
Survival				
Treatment interval	T15		• Sundi 2012[331] • Waldert 2010[332]	
Stage				
Treatment interval	T15	Waldert 2010[332]		
Prostate				
Survival				
Referral interval	T6		Neal 2007[256]	
Diagnostic interval	T8	Tørring 2013[99]		
Treatment interval	T15	O'Brien 2011[333]	• Korets 2012[334] • Sun 2012[335]	
Stage				
Patient interval	T1		Tokuda 2009[251]	
Treatment interval	T15		• Korets 2012[334] • Sun 2012[335]	

continued

TABLE 58 Summary results from previous review by Neal *et al.*[25] (*continued*)

Interval and outcome measure	Time interval[a]	Positive association	No association	Negative association
Testicular				
Survival				
Patient interval	T1	Hanson 1993[336]	Fossa 1981[337]	
Symptom onset to diagnosis	T4	• Huyghe 2007[338] • Moul 1990[339] – non-seminoma only	• Harding 1995[340] • Moul 1990[339] – seminoma only	
Symptom onset to treatment	T5	• MRC Working Party 1985[341] • Prout 1984[342]	• Dieckmann 1987[343] • Meffan 1991[344]	
Diagnostic interval	T8		Fossa 1981[337]	
Stage				
Patient interval	T1	• Hanson 1993[336] • Bosl 1981[345]	Fossa 1981[337]	
Symptom onset to diagnosis	T4	• Bosl 1981[345] • Huyghe 2007[338] • Moul 1990[339] – non-seminoma only	Harding 1995[340] Moul 1990[339] – seminoma only	
Symptom onset to treatment	T5	• Ware 1980[346] • Wishnow 1990[347] • Chilvers 1989[348]	• Dieckmann 1987[343] • Meffan 1991[344]	
Diagnostic interval	T8		• Bosl 1981[345] • Fossa 1981[337]	
Other				
Symptom onset to diagnosis and chance of complete remission	T4	Akdas 1986[349]		
Symptom onset to diagnosis and response to treatment	T4	Scher 1983[350]		
Symptom onset to treatment and relapse rate	T5		Napier 2000[351]	
Melanoma				
Survival				
Diagnostic interval	T8	• Tørring 2013[99] • Metzger 1998[352]		
Stage				
Patient interval	T1	• Temoshok 1984[353] • Montella 2002[354]	• Schmid-Wendtner 2002[355] • Carli 2003[356] • Baade 2006[357] • Helsing 1997[358]	Richard 1999[359]
Symptom onset to diagnosis	T4		• Krige 1991[360] • Baade 2006[357] • Helsing 1997[358] • Cassileth 1982[361]	

TABLE 58 Summary results from previous review by Neal *et al.*[25] (*continued*)

Interval and outcome measure	Time interval[a]	Positive association	No association	Negative association
First seen in primary care to specialist	T7	Montella 2002[354]		
Diagnostic interval	T8		• Schmid-Wendtner 2002[355] • Baade 2006[357] • Helsing 1997[358]	
Referral to specialist care to treatment	T12		Richard 1999[359]	
Non-melanoma skin				
Stage				
Patient interval	T1	Tokuda 2009[251]		
Other:				
Symptom onset to seen in specialist care and increase in tumour size	T3	Alam 2011[362]		
Symptom onset to treatment and larger lesions	T5		Renzi 2010[363]	
Gastric				
Survival				
Patient interval	T1		Lim 1974[364]	Ziliotto 1987[365]
Symptom onset to diagnosis	T4		• Arvanitakis 1992[366] • Maguire 1994[252] • Martin 1997[367] • Windham 2002[368] • Fernandez 2002[369]	Maconi 2003[370]
First (medical visit) seen in primary care to treatment	T9		Lim 1974[364]	
Treatment interval	T15		Yun 2012[236]	
Stage				
Patient interval	T1		Tokuda 2009[251]	
Diagnostic interval	T8			Haugstvedt 1991[371]
Oesophageal				
Survival				
Symptom onset to diagnosis	T4		Fernandez 2002[369]	
Stage				
Patient interval	T1		Tokuda 2009[251]	
Symptom onset to diagnosis	T4	Martin 1997[367]		
Symptom onset to treatment	T5			Wang 2008[372]

continued

TABLE 58 Summary results from previous review by Neal *et al.*[25] (*continued*)

Interval and outcome measure	Time interval[a]	Positive association	No association	Negative association
Gastro-oesophageal				
Survival				
Referral to treatment	T12		Sharpe 2010[373]	
Other				
Treatment interval and morbidity and in-hospital mortality	T15	Grotenhuis 2010[374]		
Sarcoma				
Survival:				
Symptom onset to diagnosis	T4	• Saithna 2008[375] – soft-tissue sarcoma • Nakamura 2011[376] – soft-tissue sarcoma	• Rougraff 2007[377] – soft-tissue sarcoma • Wurtz 1999[378] – osteosarcoma	
Symptom onset to treatment	T5	Ruka 1988[379] – soft-tissue sarcoma		
Stage				
Symptom onset to diagnosis	T4		Bacci 2002[380] – osteosarcoma	
Other				
Symptom onset to treatment and risk of distant metastases	T5	Nakamura 2011[376] – soft-tissue sarcoma		
Hepatocellular				
Survival				
Treatment interval	T15	Singal 2013[381]		
Stage:				
Patient interval	T1		Tokuda 2009[251]	
Renal				
Stage				
Patient interval	T1		Tokuda 2009[251]	
Symptom onset to treatment	T5			Holmang 2006[382]
Brain/central nervous system				
Stage				
Patient interval	T1	Fruchter 1981[318]	Tokuda 2009[251]	
Symptom onset to diagnosis	T4	Fruchter 1981[318]		
Diagnostic interval	T8		Fruchter 1981[318]	
Carcinoid				
Survival				
Symptom onset to diagnosis	T4		Toth-Fegel 2004[383]	

TABLE 58 Summary results from previous review by Neal *et al.*[25] (*continued*)

Interval and outcome measure	Time interval[a]	Positive association	No association	Negative association
Stage				
Symptom onset to diagnosis	T4		Toth-Fegel 2004[383]	
Multisite				
Survival				
Diagnostic interval	T8	Tørring 2013[99] – breast, lung, colorectal, prostate and melanoma combined		

MRC, Medical Research Council.
a See *Figure 2* for the time intervals.

Appendix 5 Economic decision-analytic model

Search strategy for economic models in colorectal cancer

MEDLINE

Host: Ovid.

Date range searched: 1946 to September week 3 2017.

Date searched: 29 September 2017.

Searcher: SR.

Hits: 685.

Search strategy

1. exp Economics/
2. Economics, Medical/
3. Economics, Nursing/
4. Economics, Pharmaceutical/
5. exp Economics, Hospital/
6. (economic$ or price or prices or pricing or priced or discount or discounts or discounted or discounting or ration$ or expenditure or expenditures or budget$ or afford$ or pharmacoeconomic$ or pharmaco-economic$).ti,kf.
7. exp "Fees and Charges"/
8. (fee or fees or charge$ or preference$).tw.
9. (fiscal or funding or financial or finance).tw.
10. exp "Costs and Cost Analysis"/
11. exp Health Care Costs/
12. (cost* adj2 (effective* or utilit* or benefit* or minimi* or analy* or outcome or outcomes)).ab,kf.
13. (value adj2 (money or monetary)).ti,ab,kf.
14. exp Decision Support Techniques/
15. exp Models, Economic/
16. economic model*.ab,kf.
17. (decision adj2 (tree$ or analy$ or model$)).ti,ab,kf.
18. exp Decision Theory/
19. markov.ti,ab,kf.
20. markov chains/
21. monte carlo method/
22. monte carlo.ti,ab,kf.
23. (survival adj3 analy$).tw.
24. exp Health Expenditures/
25. Uncertainty/
26. exp Budgets/
27. Decision Support Systems, Clinical/
28. (tool or tools or aid$ or model or models or checklist$ or check list$ or rule or rules or algorithm$ or equation$).tw.
29. or/1-28

30. exp Colorectal Neoplasms/
31. ((colorectal or colon* or rect* or bowel*) adj3 (cancer* or neopla* or tumor* or tumour* or carcinoma* or adenocarcinoma* or "small cell" or squamous)).tw.
32. (MCRC or CRC).tw.
33. 30 or 31 or 32
34. 29 and 33
35. Early Detection of Cancer'/
36. (predict$ or assess$ or scor$ or risk$ or validat$ or decision$ or identif$ or diagno$ or prognos$).tw.
37. (2ww or 2 week wait or two week wait or 2 week rule or two week rule).tw.
38. 35 or 36 or 37
39. 34 and 38
40. Primary Health Care/
41. exp General Practice/
42. General Practitioners/
43. (primary care or general practi$ or family practi$).tw.
44. (primary adj3 (healthcare or health care)).tw.
45. 39 or 40 or 41 or 42 or 43
46. 39 and 44

MEDLINE In-Process & Other Non-Indexed Citations

Host: Ovid.

Date range searched: not applicable.

Date searched: 29 September 2017.

Searcher: SR.

Hits: 72.

Search strategy

1. (economic$ or price or prices or pricing or priced or discount or discounts or discounted or discounting or ration$ or expenditure or expenditures or budget$ or afford$ or pharmacoeconomic$ or pharmaco-economic$).ti,kf.
2. (fee or fees or charge$ or preference$).tw.
3. (fiscal or funding or financial or finance).tw.
4. (cost* adj2 (effective* or utilit* or benefit* or minimi* or analy* or outcome or outcomes)).ab,kf.
5. (value adj2 (money or monetary)).ti,ab,kf.
6. economic model*.ab,kf.
7. (decision adj2 (tree$ or analy$ or model$)).ti,ab,kf.
8. markov.ti,ab,kf.
9. monte carlo.ti,ab,kf.
10. (survival adj3 analy$).tw.
11. (tool or tools or aid$ or model or models or checklist$ or check list$ or rule or rules or algorithm$ or equation$).tw.
12. or/1-11
13. ((colorectal or colon* or rect* or bowel*) adj3 (cancer* or neopla* or tumor* or tumour* or carcinoma* or adenocarcinoma* or "small cell" or squamous)).tw.
14. (MCRC or CRC).tw.
15. 13 or 14
16. 12 and 15

17. (predict$ or assess$ or scor$ or risk$ or validat$ or decision$ or identif$ or diagno$ or prognos$).tw.
18. (2ww or 2 week wait or two week wait or 2 week rule or two week rule).tw.
19. 17 or 18
20. 16 and 19
21. (primary care or general practi$ or family practi$).tw.
22. (primary adj3 (healthcare or health care)).tw.
23. 21 or 22
24. 20 and 23

EMBASE

Host: Ovid.

Date range searched: 1974 to 28 September 2017.

Date searched: 29 September 2017.

Searcher: SR.

Hits: 1177.

Search strategy

1. exp Economics/
2. exp Health Economics/
3. (economic$ or price or prices or pricing or priced or discount or discounts or discounted or discounting or ration$ or expenditure or expenditures or budget$ or afford$ or pharmacoeconomic$ or pharmaco-economic$).ti,kw.
4. (fee or fees or charge$ or preference$).tw
5. (fiscal or funding or financial or finance).tw.
6. Cost/
7. (cost* adj2 (effective* or utilit* or benefit* or minimi* or analy* or outcome or outcomes)).ab,kf.
8. (value adj2 (money or monetary)).ti,ab,kf.
9. Statistical Model/
10. economic model*.ab,kw.
11. markov.ti,ab,kw
12. markov chain/
13. monte carlo method/
14. monte carlo.ti,ab,kw.
15. (decision adj2 (tree$ or analy$ or model$)).ti,ab,kw.
16. Decision Theory/
17. (survival adj3 analy$).tw.
18. Budget/
19. Decision Support System/
20. (tool or tools or aid$ or model or models or checklist$ or check list$ or rule or rules or algorithm$ or equation$).tw.
21. or/1-20
22. exp Colorectal Cancer/
23. ((colorectal or colon* or rect* or bowel*) adj3 (cancer* or neopla* or tumor* or tumour* or carcinoma* or adenocarcinoma* or “small cell” or squamous)).tw.
24. (MCRC or CRC).tw.
25. 22 or 23 or 24
26. 21 and 25

27. Early Cancer Diagnosis/
28. (predict$ or assess$ or scor$ or risk$ or validat$ or decision$ or identif$ or diagno$ or prognos$).tw.
29. (2ww or 2 week wait or two week wait or 2 week rule or two week rule).tw.
30. 27 or 28 or 29
31. exp Primary Health Care/
32. General Practice/
33. General Practitioner/
34. (primary care or general practi$ or family practi$).tw.
35. (primary adj3 (healthcare or health care)).tw.
36. 31 or 32 or 33 or 34 or 35
37. 30 and 36

Health Management Information Consortium

Host: Ovid.

Date range searched: 1979 to July 2017.

Date searched: 29 September 2017.

Searcher: SR.

Hits: 154.

Search strategy

1. exp Colorectal Cancer/
2. ((colorectal or colon* or rect* or bowel*) adj3 (cancer* or neopla* or tumor* or tumour* or rcinoma* or adenocarcinoma* or "small cell" or squamous)).tw.
3. (MCRC or CRC).tw.
4. 1 or 2 or 3
5. (predict$ or assess$ or scor$ or risk$ or validat$ or decision$ or identif$ or diagno$ or prognos$).tw.
6. (2ww or 2 week wait or two week wait or 2 week rule or two week rule).tw.
7. 5 or 6
8. 4 and 7
9. exp Primary Care/
10. exp General Practice/
11. exp General Practitioners/
12. (primary care or general practi$ or family practi$).tw.
13. (primary adj3 (healthcare or health care)).tw.
14. 9 or 10 or 11 or 12 or 13
15. 8 and 14

Science Citation Index and Conference Proceedings Citation Index – Science

Host: Web of Science.

Date range searched: not applicable.

Date searched: 29 September 2017.

Searcher: SR.

Hits: 616.

Search strategy

1. TS=((pharmacoeconomic* or socioeconomics or economic* or pric* or cost* or cba or cea or cua or "health utilit*" or "value for money"))
2. TS=((colorectal or colon* or rect* or bowel*) near/2 (cancer* or neopla* or tumor* or tumour* or carcinoma* or adenocarcinoma* or "small cell" or squamous))
3. TS=(MCRC or CRC)
4. TS=(predict* or assess* or scor* or risk* or validat* or decision* or identif* or diagno* or prognos*)
5. TS=. (2ww or 2 week wait or two week wait or 2 week rule or two week rule)
6. TS=. (primary care or general practi* or family practi*)
7. TS=(primary near/2 healthcare)
8. TS=(primary near/2 health care)
9. #3 OR #2
10. #5 OR #4
11. #8 OR #7 OR #6
12. #11 AND #10 AND #9 AND #1

NHS Economic Evaluation Database and Health Technology Assessment Database

Host: Cochrane Library.

Date range searched: HTA, Issue 4 of 4, October 2016; NHS EED, 2 of 4, April 2015.

Date searched: 29 September 2017.

Searcher: SR.

Hits: HTA: 0, NHS EED; total = 0.

Search strategy

1. ((colorectal or colon* or rect* or bowel*) near/3 (cancer* or neopla* or tumor* or tumour* or carcinoma* or adenocarcinoma* or "small cell" or squamous)):ti,ab,kw
2. MeSH descriptor: [Colorectal Neoplasms] explode all trees
3. (MCRC or CRC):ti,ab,kw
4. #1 or #2 or #3
5. MeSH descriptor: [Primary Health Care] explode all trees
6. MeSH descriptor: [General Practice] explode all trees
7. MeSH descriptor: [General Practitioners] explode all trees
8. ("primary care" or "general practi*" or "family practi*"):ti,ab,kw
9. (primary near/3 (healthcare or "health care")):ti,ab,kw
10. #5 or #6 or #7 or #8 or #9
11. MeSH descriptor: [Decision Support Techniques] this term only
12. MeSH descriptor: [Decision Support Systems, Clinical] this term only
13. (tool or tools or aid* or model or models or checklist* or "check list" * or rule or rules or algorithm* or equation*):ti,ab,kw
14. #11 or #12 or #13
15. MeSH descriptor: [Early Detection of Cancer] this term only
16. (predict* or assess* or scor* or risk* or validat* or decision* or identif* or diagnos* or prognos*):ti, ab,kw
17. (2ww or "2 week wait" or "two week wait" or "2 week rule" or "two week rule"):ti,ab,kw
18. #15 or #16 or #17
19. #4 and #10 and #14 and #18

EconLit

Host: EBSCO*host*.

Date range searched: not applicable.

Date searched: 29 September 2017.

Searcher: SR.

Hits: 226.

Search strategy

1. TX ((colorectal or colon* or rect* or bowel*) N2 (cancer* or neopla* or tumor* or tumour* or carcinoma* or adenocarcinoma* or "small cell" or squamous))
2. TX (MCRC or CRC)
3. TX (predict* or assess* or scor* or risk* or validat* or decision* or identif* or diagno* or prognos*)
4. TX (2ww or 2 week wait or two week wait or 2 week rule or two week rule)
5. 1 or 2
6. 3 or 4
7. 5 and 6

Number of hits per database and in total

Database	Hits (*n*)
MEDLINE	685
MEDLINE In-Process & Other Non-Indexed Citations	72
EMBASE	1177
HMIC	154
Web of Science (SCI and CPCI-S)	616
Cochrane Library – HTA and NHS EED	0
EconLit	226
Total records	2730
Duplicates	601
Total unique records	2129

SCI, Science Citation Index.

Revised search strategy for economic models in colorectal cancer

MEDLINE

Host: Ovid.

Date range searched: 1946 to November week 4 2017.

Date searched: 11 December 2017.

Searcher: SR.

Hits: 574.

Search strategy

1	(economic$ or price or prices or pricing or priced or discount or discounts or discounted or discounting or ration$ or expenditure or expenditures or budget$ or afford$ or pharmacoeconomic$ or pharmaco-economic$).ti.
2	(fee or fees or charge$ or preference$).ti.
3	(fiscal or funding or financial or finance).ti.
4	(cost* adj2 (effective* or utilit* or benefit* or minimi* or analy* or outcome or outcomes)).ti.
5	(value adj2 (money or monetary)).ti.
6	economic model*.ti.
7	(decision adj2 (tree$ or analy$ or model$)).ti.
8	markov.ti.
9	monte carlo.ti.
10	(survival adj3 analy$).ti.
11	(tool or tools or aid$ or model or models or checklist$ or check list$ or rule or rules or algorithm$ or equation$).ti.
12	or/1-11
13	exp Colorectal Neoplasms/
14	((colorectal or colon* or rect* or bowel*) adj3 (cancer* or neopla* or tumor* or tumour* or carcinoma* or adenocarcinoma* or "small cell" or squamous)).tw.
15	(MCRC or CRC).tw.
16	or/13-15
17	"Early Detection of Cancer"/
18	(predict$ or assess$ or scor$ or risk$ or validat$ or decision$ or identif$ or diagno$ or prognos$).tw.
19	(2ww or 2 week wait or two week wait or 2 week rule or two week rule).tw.
20	or/17-19
21	12 and 16 and 20

MEDLINE In-Process & Other Non-Indexed Citations

Host: Ovid.

Date range searched: not applicable.

Date searched: 11 December 2017.

Searcher: SR.

Hits: 71.

Search strategy

1	(economic$ or price or prices or pricing or priced or discount or discounts or discounted or discounting or ration$ or expenditure or expenditures or budget$ or afford$ or pharmacoeconomic$ or pharmaco-economic$).ti.
2	(fee or fees or charge$ or preference$).ti.
3	(fiscal or funding or financial or finance).ti.
4	(cost* adj2 (effective* or utilit* or benefit* or minimi* or analy* or outcome or outcomes)).ti.
5	(value adj2 (money or monetary)).ti.

6	economic model*.ti.
7	(decision adj2 (tree$ or analy$ or model$)).ti.
8	markov.ti.
9	monte carlo.ti.
10	(survival adj3 analy$).ti.
11	(tool or tools or aid$ or model or models or checklist$ or check list$ or rule or rules or algorithm$ or equation$).ti.
12	or/1-11
13	((colorectal or colon* or rect* or bowel*) adj3 (cancer* or neopla* or tumor* or tumour* or carcinoma* or adenocarcinoma* or "small cell" or squamous)).tw.
14	(MCRC or CRC).tw.
15	13 or 14
16	(predict$ or assess$ or scor$ or risk$ or validat$ or decision$ or identif$ or diagno$ or prognos$).tw.
17	(2ww or 2 week wait or two week wait or 2 week rule or two week rule).tw.
18	16 or 17
19	12 and 15 and 18

EMBASE

Host: Ovid.

Date range searched: 1974 to 9 December 2017.

Date searched: 11 December 2017.

Searcher: SR.

Hits: 658.

Search strategy

1	(economic$ or price or prices or pricing or priced or discount or discounts or discounted or discounting or ration$ or expenditure or expenditures or budget$ or afford$ or pharmacoeconomic$ or pharmaco-economic$).ti.
2	(fee or fees or charge$ or preference$).ti.
3	(fiscal or funding or financial or finance).ti.
4	(cost* adj2 (effective* or utilit* or benefit* or minimi* or analy* or outcome or outcomes)).ti.
5	(value adj2 (money or monetary)).ti.
6	economic model*.ti.
7	(decision adj2 (tree$ or analy$ or model$)).ti.
8	markov.ti.
9	monte carlo.ti.
10	(survival adj3 analy$).ti.
11	(tool or tools or aid$ or model or models or checklist$ or check list$ or rule or rules or algorithm$ or equation$).ti.
12	or/1-11
13	exp colorectal cancer/
14	((colorectal or colon* or rect* or bowel*) adj3 (cancer* or neopla* or tumor* or tumour* or carcinoma* or adenocarcinoma* or "small cell" or squamous)).tw.

15	(MCRC or CRC).tw.
16	or/13-15
17	early cancer diagnosis/
18	(predict$ or assess$ or scor$ or risk$ or validat$ or decision$ or identif$ or diagno$ or prognos$).tw.
19	(2ww or 2 week wait or two week wait or 2 week rule or two week rule).tw.
20	or/17-19
21	12 and 16 and 20

Health Management Information Consortium

Host: Ovid.

Date range searched: 1979 to September 2017.

Date searched: 11 December 2017.

Searcher: SR.

Hits: 32.

Search strategy

1. exp Colorectal Cancer/
2. ((colorectal or colon* or rect* or bowel*) adj3 (cancer* or neopla* or tumor* or tumour* or carcinoma* or adenocarcinoma* or "small cell" or squamous)).tw.
3. (MCRC or CRC).tw.
4. 1 or 2 or 3
5. (predict$ or assess$ or scor$ or risk$ or validat$ or decision$ or identif$ or diagno$ or prognos$).tw.
6. (2ww or 2 week wait or two week wait or 2 week rule or two week rule).tw.
7. 5 or 6
8. 4 and 7

NHS Economic Evaluation Database and Health Technology Assessment Database

Host: Cochrane Library.

Date range searched: Health Technology Assessment, issue 4 of 4, October 2016; NHS EED, 2 of 4, April 2015.

Date searched: 11 December 2017.

Searcher: SR.

Hits: HTA: 0; NHS EED: 0; total = 0.

Search strategy

1. ((colorectal or colon* or rect* or bowel*) near/3 (cancer* or neopla* or tumor* or tumour* or carcinoma* or adenocarcinoma* or "small cell" or squamous)):ti,ab,kw
2. MeSH descriptor: [Colorectal Neoplasms] explode all trees
3. (MCRC or CRC):ti,ab,kw
4. #1 or #2 or #3

5. MeSH descriptor: [Early Detection of Cancer] this term only
6. (predict* or assess* or scor* or risk* or validat* or decision* or identif* or diagnos* or prognos*):ti,ab,kw
7. (2ww or "2 week wait" or "two week wait" or "2 week rule" or "two week rule"):ti,ab,kw
8. #5 or #6 or #7
9. #4 and #8

EconLit

Host: EBSCO*host*.

Date range searched: not applicable.

Date searched: 11 December 2017.

Searcher: SR.

Hits: 1.

Search strategy

1. TX ((colorectal or colon* or rect* or bowel*) N2 (cancer* or neopla* or tumor* or tumour* or carcinoma* or adenocarcinoma* or "small cell" or squamous))
2. TX (MCRC or CRC)
3. TX (predict* or assess* or scor* or risk* or validat* or decision* or identif* or diagno* or prognos*)
4. TX (2ww or 2 week wait or two week wait or 2 week rule or two week rule)
5. 1 or 2
6. 3 or 4
7. 5 and 6

Number of hits per database and in total

Database	Hits (*n*)
MEDLINE	574
MEDLINE In-Process & Other Non-Indexed Citations	71
EMBASE	658
HMIC	32
Cochrane Library – HTA and NHS EED	0
EconLit	1
Total records	1336
Duplicates	564
Total unique records	772

Updated searches, 27 September 2018

Database	Hits (*n*)
MEDLINE	249
MEDLINE In-Process & Other Non-Indexed Citations	28
EMBASE	351
HMIC	0

Database	Hits (*n*)
Cochrane Library – HTA and NHS EED	1
EconLit	0
Total records	629
Duplicates	83
Total unique records	546

American Joint Committee on Cancer and Dukes' classification systems

The decision models in the literature are based on two different classification systems of CRC stages. These are the AJCC and Dukes.

The American Joint Committee on Cancer staging system

The AJCC's staging system[384] classifies cancer following the TNM system, in which the first consideration is the size and extent of the main tumour (or primary tumour), the second refers to the number of nearby lymph nodes with cancer, and the third one refers whether or not the cancer has metastasised, if the cancer has spread from the primary tumour to other parts of the body. The TNM system is the most widely used cancer staging system (see www.cancer.gov/about-cancer/diagnosis-staging/staging; accessed January 2019).

Primary tumours can be graded in six levels:

1. T0, no evidence of primary tumour
2. Tis, carcinoma in situ: intraepithelial or invasion of the lamina propria
3. T1, the tumour is confined to the submucosa
4. T2, the tumour has grown into (but not through) the muscularis propria
5. T3, the tumour has grown into (but not through) the serosa
6. T4, the tumour directly invades other organs or structures, and/or perforates visceral peritoneum.

The levels for lymph nodes are as follows:

- N0, no regional nodes involved
- N1, 1–3 regional nodes involved
- N2, four or more regional nodes involved.

Finally, metastasis can be:

- M0: no distant metastasis
- M1: distant metastasis present.

Tumour, number, metastasis combinations were grouped into five stages:

1. Stage 0, cancer is defined by Tis N0 M0, called carcinoma in situ, is related to the presence of abnormal cells, but they have not spread to nearby tissue.
2. Stage I, a cancer defined by T1 N0 M0 or T2 N0 M0.
3. Stage II, a cancer defined by T3–4 N0 M0.
4. Stage III, represented by cancer defined by any T1–2, N1, M0, any T3–4, N1 M0, any T N2 M0.
5. Stage IV, represented by a cancer with any T, any N, M1, where the cancer has spread to distant parts of the body.

The Dukes' staging system

The other staging system was initially defined by Dukes in 1932,[385] but this has been revised several times. This system has now largely been replaced by the TNM staging system, but it is currently used in many of the English and UK economic models. This system has four stages:

1. Dukes' A: invasion into but not through the bowel wall (90% chance of 5-year survival)
2. Dukes' B: invasion through the bowel wall but not involving lymph nodes (70% chance of 5-year survival)
3. Dukes' C: involvement of lymph nodes (30% chance of 5-year survival)
4. Dukes' D: widespread metastases.

Comparison of the American Joint Committee on Cancer and Dukes' staging systems

A conversion of the Dukes' stages into TNM is represented in *Table 59*.

Excluded studies of full-text articles assessed for eligibility

Thirteen full texts were excluded because the models were not based on the Dukes' or AJCC classification staging (9/13), the models were screening models without a full disease progression model component (3/13) or the study was a costing analysis for a screening programme (1/13). Further details are chronologically summarised in this section.

Bolin *et al.*[386] compared the costs of different strategies for CRC screening, including colonoscopy, to provide a decision on early diagnosis for CRC in Australia. This model was based on a model developed by the Office of Technology Assessment (OTA).[246] Both models were excluded because the staging was grouped into A + B and C + D, and did not use the Dukes' full staging classification. The model by Byers and Gorsky[387] was also excluded because it too used the OTA model, which considers only two stages based on an incomplete Dukes' classification system.

Neilson and Whynes[388] reported the development of a simulation model for assessing the cost-effectiveness of CRC in the UK. The model was not included as it considers only two health states for CRC (early- and late-detected CRC), which does not correspond with the Dukes' staging classification.

Gow's[389] study was excluded on the basis of being a costing study of a programme in Australia for screening CRC.

TABLE 59 Conversion of Dukes' stages into TNM

Stage	Dukes'	TNM
Tumour invasion confined to the mucosa	A	Tis, N0
Tumour invasion limited to the submucosa, no lymph node involvement	A	T1, N0
Tumour invasion limited to the submucosa, lymph node involvement	C	T1, N1–2
Limited tumour invasion into the muscle layer, no lymph node involvement	A	T2, N0
Limited tumour invasion into the muscle layer, lymph node involvement	C	T2, N1–2
During the whole muscle layer tumour involvement, no lymph node involvement	B	T3, N0
During the whole muscle layer tumour involvement, lymph node involvement	C	T3, N1–2
Tumours have kept the neighbouring organs, no lymph node involvement	B	T4, N0
Tumours have kept the neighbouring organs, lymph node involvement	C	T4, N1–2
Other factors notwithstanding distant metastases	D	T1–4, N0–2, M1

Khandker *et al.*[390] evaluated the guidelines of the American Gastroenterological Association for CRC screening and surveillance programmes, using a cost-effectiveness framework for people at average risk and increased risk in the USA. This study was not included because the stages of CRC were dependent on anatomical extent (local, regional and distant) that only conceptually parallels with three stages (A, B and C) of the Dukes' staging classification.

The study by Sonnenberg *et al.*[391] compared the cost-effectiveness of faecal occult blood testing, flexible sigmoidoscopy and colonoscopy as strategies for screening CRC. The study was excluded because the Markov models were for the screening strategies only and did not include a full disease model component.

Reyes *et al.*[392] was excluded because the authors were evaluating the clinical effectiveness and incremental cost-effectiveness of four strategies to detect hereditary non-polyposis colorectal carcinoma (HNPCC) gene carriers in individuals with CRC and the authors excluded a full disease progression model.

Kievit *et al.*[393] determined the clinical effectiveness, efficiency and feasibility of a new strategy based on microsatellite instability analysis for the detection of HNPCC in the Netherlands. This study was excluded because the Markov model included only one health state for CRC, without considering the different stages of the disease.

Telford *et al.*[394] estimated the incremental cost-effectiveness of 10 strategies for CRC screening in Canada. The study was not included because the Dukes' staging or other classification system that was part of the inclusion criteria was not used.

The study by Allameh *et al.*[395] was excluded because this was another screening cost-effectiveness model that did not include a whole-disease progression model.

The study by Dan *et al.*[396] was excluded because the cost-effectiveness analysis of single endoscopic examination screening within a Markov framework was based on the risk for the population of Singapore and based on a single health state (progression to CRC).

The methodological paper by Goh *et al.*[341] was excluded because the study aimed to compute the maximal and minimal values for the discounted value of the Markov chain for the transition parameters of CRC disease in the USA. The staging used in this model was localised, regional and distant cancer sites, which differs from the Dukes' or other system classification specified in the inclusion criteria.

Finally, a 2017 study by Atkin *et al.*[397] was also excluded because the authors considered five mutually exclusive health states (no adenomas, adenomas, preclinical CRC, diagnosed CRC and dead), and thus omitted any type of staging classification.

Sensitivity analysis, parameter specification and incremental cost-effectiveness ratio variation

TABLE 60 Sensitivity analysis, parameter specification and ICER variation

Study	Type of analysis	Parameters affecting ICER	ICER variation
Allen 2005[144]	• Univariate sensitivity analysis (USA) • Two-way sensitivity analysis	• Prevalence of IBD • Time at which 90% of IBD cases are diagnosed • Risk of perforation with colonoscopy • Risk of haemorrhage with colonoscopy • Sensitivity of FS for IBD • Age at entry in the screening model • Polyp prevalence • Cost of colonoscopy • Cost of FS	• The increase in IBD prevalence (from 2.2% to 20%) increases the ICER of colonoscopy, compared with FS • A prolonged time to diagnosis of IBD (from 0.5 to 3.5 years) increase the ICER of colonoscopy, compared with FS • The increase in risk of perforation of colon during colonoscopy procedures (from 0.001% to 2.14%) increases the ICER of colonoscopy, compared with FS • The increase in risk of haemorrhage from colonoscopy procedures (from 0.24% to 4.6%) increases the ICER of colonoscopy, compared with FS • The increase in FS sensitivity (from 25% to 80%) reduces the ICER of colonoscopy, compared with FS • The increase of FS sensitivity for rectosigmoid polyps (from 75% to 98%) increases the ICER of colonoscopy, compared with FS • The increase of age of entry (from 40 to 80 years) increases the ICER of all invasive procedures (colonoscopy, FS, FS + ACBE) • The decrease of polyp prevalence (from 27% to 7%) increases the ICER of all invasive procedures (colonoscopy, FS, FS + ACBE) • The increase in the cost of colonoscopy (from US$200 to US$2200) increases the ICER of colonoscopy, compared with watchful waiting, FS and FS + ACBE • The increase in the cost of FS (from US$100 to US$844) increases the ICER of FS, compared with colonoscopy

TABLE 60 Sensitivity analysis, parameter specification and ICER variation (*continued*)

Study	Type of analysis	Parameters affecting ICER	ICER variation
Tappenden 2007[145]	• One-way sensitivity analysis • Probabilistic sensitivity analysis	• Undiscounted costs and effects • 40% of individuals never participate in screening • 60% compliance with follow-up colonoscopy • Doubled CRC treatment costs • 20% lower sensitivity for FOBT and screening • 10% FOBT sensitivity for all adenomas • 5% FOBT sensitivity for low-risk adenomas and 10% FOBT sensitivity for high-risk adenomas • Double adenoma recurrence rates following polypectomy • Utility for all cancer states = 0.50 • FS cost = colonoscopy cost • Double cost of FOBT • Best-case scenario for calibrated natural history and sensitivity parameters • Worst-case scenario for calibrated natural history and sensitivity parameters	• The undiscounted costs and effects reduces the ICER of the strategies • The adoption of 40% of missed participants' rate to FOBT screening does not affect the ICER • The adoption of 60% of compliance rate with follow-up of colonoscopy increases the ICER of FOBT screening strategies • The increase of CRC treatment costs reduces the ICER of screening strategies • The reduction of FOBT sensitivity and FS screening increases the ICER of both FOBT and FS screening • The adoption of 10% of FOBT sensitivity for all adenomas reduces the ICER of FOBT screening strategies • The adoption of 5% FOBT sensitivity for low-risk adenomas and 10% of FOBT sensitivity for high-risk adenomas reduces the ICER of FOBT screening strategies • The increase (doubled) in an adenoma's recurrence rate following a polypectomy increases the ICER • The adoption of the same utility value (0.50) for all cancer states increases the ICER • The adoption of the same cost of colonoscopy for FS increases the ICER • The increase of FOBT (doubled) increases the ICER for FOBT screening strategies • The adoption of best-case scenario for calibrated natural history and sensitivity parameters reduces the ICER • The adoption of worst-case scenario for calibrated natural history and sensitivity parameters increases the ICER
Tsoi 2008[146]	• One-way sensitivity analysis • Two-way sensitivity analysis	• Cost of colonoscopy • Compliance rate of screening tests • FOBT sensitivity • FOBT specificity • Treatment costs for CRC • Discount rates	• The increase in the cost of colonoscopy (from US$100 to US$1000) increases the ICER for all strategies • The increase in the compliance rate of screening tests reduces the ICER • The increase in FOBT sensitivity (from 10% to 100%) reduces the ICER for FOBT strategy

continued

TABLE 60 Sensitivity analysis, parameter specification and ICER variation (*continued*)

Study	Type of analysis	Parameters affecting ICER	ICER variation
			• The increase of FOBT specificity (20%, 50% and 80%) reduces the ICER for FOBT strategy • The increase of treatment costs for CRC reduces the ICER for all strategies • The adoption of discount rate for cost and clinical effectiveness increases the ICER for all strategies
Zauber 2008[147]	N/A	N/A	N/A
Heitman 2010[148]	• USA • Probabilistic sensitivity analysis	• Increase in FIT cost • Decrease of cancer stage III and IV costs • Biennial FIT screening • Initial adherence of FIT, FOBT, CTC and colonoscopy • Decrease subsequent of adherence rates for FITs and FOBTs • Administrative cost	• The increase in FIT cost (50%) increases the ICER of FIT screening strategies • The reduction of cancer stages III and IV costs increases the ICER for all strategies • The adoption of biennial screening reduces the ICER of FIT screening strategies • The adoption of initial adherence rate for FIT, FOBT, CTC and colonoscopy set at 60%, 50%, 40% and 30% increases the ICER for all strategies • The subsequent decrease of adherence rates for FITs and FOBTs (from 63% to 40%) reduces the ICER of FITs and increases the ICER of colonoscopy strategies. The reduction from 63% to 20% reduces the ICER of FITs and colonoscopy strategies with respect to no screening strategies • The adoption of administrative costs added for all screening tests (US$10) increased the ICER of FIT-high strategy and reduces the ICER of FIT-mid strategy. The increase of US$50 reduced the ICER of colonoscopy
Lee 2010[149]	• USA • Probabilistic sensitivity analysis	• Screening uptake • Polyp detection sensitivity • Lifetime treatment costs • Screening costs • Adverse event probability • Adverse event cost • Cancer detection sensitivity	• The decrease of the uptake rates of the two screening strategies (from 80% to 40%) decreases the ICER for CTC, compared with FOBT • The polyp detection sensitivity of CTC and FOBT (from 75% to 56% for low-risk CTC, from 96% to 84% for high-risk CTC, and from 10% to 0% for all FOBT) reduced the ICER for CTC, compared with FOBT

TABLE 60 Sensitivity analysis, parameter specification and ICER variation (*continued*)

Study	Type of analysis	Parameters affecting ICER	ICER variation
			• The increase of lifetime treatment costs of Dukes' stage C and D cancer (20% higher) the ICER of CTC decrease when compared with FOBT • The decrease of screening costs increases the ICER of CTC when compared with FOBT • The increase of adverse event probability increases the ICER of CTC when compared with FOBT • The increase of adverse event cost increases the ICER of CTC when compared with FOBT • The cancer detection sensitivity of CTC (from 96% to 84%) and FOBT (from 50% to 30%) reduces the ICER of CTC when compared with FOBT
Knudsen 2012[150]	USA	• Rescreening test characteristics • Adherence rates • Cost of colonoscopy	• The adoption of use of CTC rescreening after 5 years from colonoscopy strategy decreases the ICER, compared with the use of colonoscopy 10 years after the previous colonoscopy • The increase of adherence rates decreases the ICER of all strategies • The decrease of adherence rates increases the ICER of all strategies • The increase in the cost of colonoscopy (one half) decreases the ICER of all rescreening strategies (including continuing colonoscopy)
Sharp 2012[151]	• USA • Probabilistic sensitivity analysis	• FIT sensitivity • Colonoscopy sensitivity • Cost of FIT • Cost of colonoscopy • Lifetime costs of treating CRC • Discount rate	• The increase of FIT sensitivity increases the ICER • The increase in colonoscopy sensitivity increases the ICER, compared with no screening strategy • The increase in costs of FIT increases the ICER, compared with no screening strategy • The increase in the cost of colonoscopy increases the ICER for all screening strategies, compared with no screening strategy • The increase in the cost of CRC treatment increases the ICER for all screening strategies, compared with no screening strategy

continued

TABLE 60 Sensitivity analysis, parameter specification and ICER variation (*continued*)

Study	Type of analysis	Parameters affecting ICER	ICER variation
			• The adoption of undiscounted costs and benefits decreases the ICER for all strategies
Whyte 2012[152]	• One-way sensitivity analysis • Probabilistic sensitivity analysis	• Screening uptake • Endoscopic test costs • FIT sensitivity and specificity	• The increase of uptake (from 54% to 70%) decreases the ICER • The increase of endoscopic test costs increases the ICER • The adoption of FIT with higher thresholds (150 and 200 ng Hb/ml) decreases the ICER
Goede 2013[153]	Digital Subtraction Angiography	• FIT costs • Colonoscopy capacity	• The decrease of 50% of FIT cost decreases the ICER decreased of the two-sample strategy, compared with the one-sample strategy • The capacity limit of five colonoscopies per 1000 individuals per year decreases the ICER of two-sample FIT strategy, compared with one-sample FIT strategy
Gomes 2013[154]	• USA • Probabilistic sensitivity analysis	• Polyp detection sensitivity of imaging test (CTC, OC) for polyps of < 5 mm • Polyp detection sensitivity of imaging test (CTC, OC) for polyps of between 6 and 9 mm • Polyp detection sensitivity of imaging test (CTC, OC) for polyps of > 10 mm • CRC detection sensitivity • Discount rate • Cost of OC test • Cost of CTC test • Lifetime treatment costs • Utilities • Utility values	• The increase of OC sensitivity for detection of polyps of < 5 mm (from 0% to 96%) decreases the INB, whereas an increase of CTC sensitivity for detection of polyps of < 5 mm (from 0% to 90%) increases the INB for CTC, compared with OC • The increase of OC sensitivity for detection of polyps of between 6 and 9 mm (from 60% to 100%) decreases the INB, whereas an increase of CTC sensitivity for detection of polyps of between 6 and 9 mm (from 60% to 100%) increases the INB for CTC, compared with OC • The increase of OC sensitivity for detection of polyps of > 10 mm (from 80% to 100%) decreases the INB, whereas an increase of CTC sensitivity for detection of polyps of > 10 mm (from 80% to 100%) increases the INB for CTC, compared with OC • The increase of OC sensitivity for CRC detection (from 80% to 100%) decreases the INB, whereas an increase of CTC sensitivity for CRC detection (from 80% to 100%) increases the INB for CTC, compared with OC • The increase of the discount rate (from 0% to 5%) increases the INB for CTC, compared with OC

TABLE 60 Sensitivity analysis, parameter specification and ICER variation (*continued*)

Study	Type of analysis	Parameters affecting ICER	ICER variation
			• The increase of OC (30% higher) increases the INB for CTC, compared with OC • The increase of CTC (30% higher) decreases the INB for CTC, compared with OC • The increase of lifetime treatment costs (30% higher) increases the INB for CTC, compared with OC • The adoption of 0.5 as utility value for all health states increases the INB for CTC, compared with OC • The adoption of 0.8 as utility value for all health states decreases the INB for CTC, compared with OC
Tappenden 2013[155]	No sensitivity analysis	The authors provided constrained maximisation analysis	N/A
Whyte 2014[156]	• One-way sensitivity analysis • Two-way sensitivity analysis • Probabilistic sensitivity analysis	• Magnitude, duration of the increase of presentation rates • Increase at Dukes' stage presentation	• An increase of 1 month of the duration of change in symptomatic presentation rate and an increase of 5% of the magnitude of change in symptomatic presentation rate increases the ICER • An increase of 1 month in the duration of change in symptomatic presentation rate and an increase of 20% in the magnitude of change in symptomatic presentation rate reduces the ICER • An increase of 3 or 6 months in the duration of change in symptomatic presentation rate and an increase of 5%, 10% or 20% of the magnitude of change in symptomatic presentation rate reduces the ICER • An increase of 1 month of the duration of increase in symptomatic presentation rate and an increase of 5%, 10% or 20% of the magnitude of change in symptomatic presentation rate increases the ICER • An increase of 3 months of the duration of increase in symptomatic presentation rate and an increase of 5% of the magnitude of change in symptomatic presentation rate increases the ICER

continued

TABLE 60 Sensitivity analysis, parameter specification and ICER variation (*continued*)

Study	Type of analysis	Parameters affecting ICER	ICER variation
			• An increase of 3 months of the duration of increase in symptomatic presentation rate and an increase of 10% or 20% of the magnitude of change in symptomatic presentation rate reduces the ICER • An increase of 6 months of the duration of increase in symptomatic presentation rate and an increase of 5%, 10% or 20% of the magnitude of change in symptomatic presentation rate reduces the ICER • The increase of presentation rate at each CRC Dukes' stage (5%, 10%, 20%) decreases the ICER
Cantor 2015[157]	• One-way sensitivity analysis • Two-way sensitivity analysis	• Percentage change in individual's behaviour • Cost of decision aid • Cost of patient time to use the decision aid	• The increase of percentage of people who changed from no screening to screening decreases the ICER • The decrease of the decision aid cost decreases the ICER • The increase of cost of patient time increases the ICER
Pil 2016[158]	• One-way sensitivity analysis • Probabilistic sensitivity analysis	• Sensitivity of FIT for high-risk polyps • Natural progression of CRC • Specificity of FIT • Prevalence of unidentified high-risk polyps • Adherence to colonoscopy after referral • Sensitivity of colonoscopy for high-risk polyps	• The increase of sensitivity of the FIT for high-risk polyps (from 42% to 214%) decreases the ICER • The increase of the natural progression of CRC (from 70% to 130%) decreases the ICER • The increase of the specificity of the FIT (from 80% to 103%) decreases the ICER • The increase of the prevalence of unidentified high-risk polyps (from 70% to 130%) decreases the ICER • The increase of the adherence to colonoscopy after referral decreases the ICER • The increase of sensitivity of colonoscopy for high-risk polyps decreases the ICER
Wong 2016[159]	• One-way sensitivity analysis • Probabilistic sensitivity analysis	• FS compliance rate • FS specificity • Colonoscopy cost • FS cost	• The decrease of FS compliance rate (from 100% to 0%), increases the ICER • The increase of the specificity of FS (from 60% to 90%) decreases the ICER

TABLE 60 Sensitivity analysis, parameter specification and ICER variation (*continued*)

Study	Type of analysis	Parameters affecting ICER	ICER variation
			• The increase of cost of colonoscopy (from US$100 to US$450) increases the ICER of strategies for which colonoscopy is included • The increase of cost of FS increases the ICER of the strategies for which FS is included
Coldman 2017[160]	DSA	• FIT threshold • Test positivity rates (FIT sensitivity rate in subjects with CRC or adenoma, and the false-positive rates in subjects without neoplasia)	• Evaluation of clinical effectiveness only in the sensitivity analysis
Westwood 2017[161]	• Scenario analysis • Probabilistic sensitivity analysis	• Mortality rate given by colonoscopy • Threshold for FIT assay • Prevalence of CRC • Cost test • CRC mortality progression • Probability of persisting in symptoms after negative test result • High rate of mortality associated with colonoscopy • Exclusion of adverse events associated with colonoscopy • Second FIT/gFOBT for patients with a negative test result when symptoms persist	• The increase in the mortality rate given by colonoscopy increases the ICER • The adoption of a threshold of 2 µg Hb/g faeces decreases the ICER • The adoption of 3% and 5.4% for CRC prevalence decreases the ICER for FIT, compared with gFOBT • The exclusion of CRC progression mortality reduces the ICER • The inclusion of double probability of persisting in symptoms after a negative test result decreases the ICER • The inclusion of half probability of persisting in symptoms after a negative test result decreases the ICER • The inclusion of a high rate of mortality associated with colonoscopy decreases the ICER • The exclusion of adverse events associated with colonoscopy decreases the ICER • The adoption of a second FIT/gFOBT for patients with a negative test result when symptoms persist increases the ICER

ACBE, air-contrast barium enema; FS, flexible sigmoidoscopy; IBD, inflammatory bowel disease; INB, incremental net benefit; N/A, not applicable; OC, optical colonoscopy.

Critical appraisal of model-based economic evaluations

TABLE 61 Critical appraisal of model-based economic evaluation, Philips *et al.*[162]

Question	Allen *et al.*[144]	Tappenden *et al.*[145]	Tsoi *et al.*[146]	Zauber *et al.*[147]	Heitman *et al.*[148]	Lee *et al.*[149]	Knudsen *et al.*[150]	Sharp *et al.*[151]	Whyte *et al.*[152]	Goede *et al.*[153]	Gomes *et al.*[154]	Tappenden *et al.*[155]	Whyte and Harnan[156]	Cantor *et al.*[157]	Pil *et al.*[158]	Wong *et al.*[159]	Coldman *et al.*[160]	Westwood *et al.*[161]
Structure																		
S1: statement of decision problem/objective																		
Is there a clear statement of the decision problem?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Is the objective of the evaluation and model specified and consistent with the stated decision problem?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Is the primary decision-maker specified?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
S2: statement of scope/perspective																		
Is the perspective of the model stated clearly?	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No	Yes	Yes
Are the model inputs consistent with the stated perspective?	Yes	Yes	NR	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Has the scope of the model been stated and justified?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Are the outcomes of the model consistent with the perspective, scope and overall objective of the model?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
S3: rationale for model structure																		
Has the evidence regarding the model structure been described?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Is the structure of the model consistent with a coherent theory of the health condition under evaluation?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Question	Allen *et al.*[144]	Tappenden *et al.*[145]	Tsoi *et al.*[146]	Zauber *et al.*[147]	Heitman *et al.*[148]	Lee *et al.*[149]	Knudsen *et al.*[150]	Sharp *et al.*[151]	Whyte *et al.*[152]	Goede *et al.*[153]	Gomes *et al.*[154]	Tappenden *et al.*[155]	Whyte and Harnan[156]	Cantor *et al.*[157]	Pil *et al.*[158]	Wong *et al.*[159]	Coldman *et al.*[160]	Westwood *et al.*[161]
Have any competing theories regarding model structure been considered?	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
Are the sources of data used to develop the structure of the model specified?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Are the causal relationships described by the model structure justified appropriately?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
S4: structural assumptions																		
Are the structural assumptions transparent and justified?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Are the structural assumptions reasonable given the overall objective, perspective and scope of the model?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
S5: strategies/comparators																		
Is there a clear definition of the option under evaluation?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Have all feasible and practical options been evaluated?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Is there justification for the exclusion of feasible options?	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
S6: model type																		
Is the chosen model type appropriate given the decision problem and specified causal relationship with the model?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
S7: time horizon																		
Is the time horizon of the model sufficient to reflect all important differences between options?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Is the time horizon of the model, and the duration of treatment and treatment effect described and justified?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

continued

TABLE 61 Critical appraisal of model-based economic evaluation, Philips *et al.*[162] (*continued*)

Question	Allen *et al.*[144]	Tappenden *et al.*[145]	Tsoi *et al.*[146]	Zauber *et al.*[147]	Heitman *et al.*[148]	Lee *et al.*[149]	Knudsen *et al.*[150]	Sharp *et al.*[151]	Whyte *et al.*[152]	Goede *et al.*[153]	Gomes *et al.*[154]	Tappenden *et al.*[155]	Whyte and Harnan[156]	Cantor *et al.*[157]	Pil *et al.*[158]	Wong *et al.*[159]	Coldman *et al.*[160]	Westwood *et al.*[161]
S8: disease states/pathways																		
Do the disease states (state-transition model) or the pathways (decision tree model) reflect the underlying biological process of the disease in question and the impact of the interventions?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
S9: cycle length																		
Is the cycle length defined and justified in terms of the natural history of the disease?	No	Yes	No	Yes	Yes	No	No	Yes	Yes	No	No	No	No	N/A	No	No	No	Yes
Data																		
D1: data identification																		
Are the data identification methods transparent and appropriate given the objectives of the model?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Where choices have been made between data sources, are these justified appropriately?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Has particular attention been paid to identifying data for the important parameters in the model?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Has the process of selecting key parameters been justified and systematic methods used to identify the most appropriate data?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Has the quality of the data been assessed appropriately?	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes
Where expert opinion has been used, are the methods described and justified?	No	Yes	N/A	N/A	N/A	N/A	No	Yes	Yes	N/A	N/A	Yes	No	N/A	Yes	N/A	Yes	Yes
D2: pre-model data analysis																		
Are the pre-model data analysis methodology based on justifiable statistical and epidemiological techniques?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Question	Allen *et al.*[144]	Tappenden *et al.*[145]	Tsoi *et al.*[146]	Zauber *et al.*[147]	Heitman *et al.*[148]	Lee *et al.*[149]	Knudsen *et al.*[150]	Sharp *et al.*[151]	Whyte *et al.*[152]	Goede *et al.*[153]	Gomes *et al.*[154]	Tappenden *et al.*[155]	Whyte and Harnan[156]	Cantor *et al.*[157]	Pil *et al.*[158]	Wong *et al.*[159]	Coldman *et al.*[160]	Westwood *et al.*[161]
D2a: baseline data																		
Is the choice of baseline date described and justified?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	N/A	Yes	Yes	Yes	Yes
Are transition probabilities calculated appropriately?	NR	Yes	Yes	Yes	Yes	N/A	N/A	Yes	Yes	N/A	Yes	Yes	Yes	N/A	Yes	Yes	Yes	Yes
Has a half cycle correction been applied to both cost and outcome?	NR	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
If not, has this omission been justified?	NR	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
D2b: treatment effects																		
If relative treatment effects have been derived from trial data, have they been synthesised using appropriate techniques?	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Have the methods and assumptions used to extrapolate short-term results to final outcomes been documented and justified?	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Have alternative assumptions been explored through sensitivity analysis?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Have assumptions regarding the continuing effect of treatment once treatment is complete been documented and justified?	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	No	N/A	N/A	N/A	N/A	N/A	No	N/A
D2c: quality-of-life weights (utilities)																		
Are the utilities incorporated into the model appropriate?	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
Is the source for the utility weights referenced?	Yes	Yes	Yes	Yes	Yes	No	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
Are the methods of derivation for the utility weights justified?	No	No	No	No	No	No	No	No	No	No	Yes	No	No	No	Yes	Yes	No	Yes

continued

TABLE 61 Critical appraisal of model-based economic evaluation, Philips *et al.*[162] (*continued*)

Question	Allen *et al.*[144]	Tappenden *et al.*[145]	Tsoi *et al.*[146]	Zauber *et al.*[147]	Heitman *et al.*[148]	Lee *et al.*[149]	Knudsen *et al.*[150]	Sharp *et al.*[151]	Whyte *et al.*[152]	Goede *et al.*[153]	Gomes *et al.*[154]	Tappenden *et al.*[155]	Whyte and Harnan[156]	Cantor *et al.*[157]	Pil *et al.*[158]	Wong *et al.*[159]	Coldman *et al.*[160]	Westwood *et al.*[161]
D3: data incorporation																		
Have all the data incorporated into the model been described and referenced in sufficient detail?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes
Has the use of mutually inconsistent data been justified (i.e. are assumptions and choices appropriate?)	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	N/A	Yes	Yes	Yes	N/A	Yes	No	Yes
Is the process of data incorporation transparent?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	No	Yes	Yes	No	No	Yes
If the data have been incorporated as distributions, has the choice of distribution for each parameter been described and justified?	No	Yes	N/A	N/A	N/A	N/A	N/A	Yes	Yes	N/A	Yes	No	N/A	No	Yes	No	No	Yes
If data have been incorporated as distributions, is it clear that second-order uncertainty is reflected?	No	Yes	N/A	N/A	N/A	N/A	N/A	Yes	Yes	N/A	No	Yes	N/A	No	Yes	No	No	Yes
D4: assessment of uncertainty																		
Have the four principal types of uncertainty been addressed?	No	Yes	No	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	No	Yes	No	No	No
If not, has the omission of particular forms of uncertainty been justified?	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
Have methodological uncertainties been addressed by running alternative versions of the model with different methodological assumptions?	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	No	No	No
Is there evidence that structural uncertainties have been addressed via sensitivity analysis?	No	Yes	No	No	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	No	No	No
D4c: heterogeneity																		
Has heterogeneity been dealt with by running the model separately for different subgroups?	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Question	Allen *et al.*[144]	Tappenden *et al.*[145]	Tsoi *et al.*[146]	Zauber *et al.*[147]	Heitman *et al.*[148]	Lee *et al.*[149]	Knudsen *et al.*[150]	Sharp *et al.*[151]	Whyte *et al.*[152]	Goede *et al.*[153]	Gomes *et al.*[154]	Tappenden *et al.*[155]	Whyte and Harnan[156]	Cantor *et al.*[157]	Pil *et al.*[158]	Wong *et al.*[159]	Coldman *et al.*[160]	Westwood *et al.*[161]
D4d: parameter																		
Are the methods of assessment of parameter uncertainty appropriate?	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	Yes	No	No	Yes	Yes	Yes	Yes	No	No	Yes
Has probabilistic sensitivity analysis been done; if not, has this been justified?	No	N/A	No	No	N/A	N/A	No	N/A	N/A	Yes	N/A	N/A	N/A	No	N/A	N/A	No	Yes
If the data are incorporated as point estimates, are the ranges used for sensitivity analysis stated clearly and justified?	Yes	N/A	Yes	N/A	Yes	Yes	N/A	Yes	Yes	Yes	Yes	Yes	Yes	N/A	No	N/A	N/A	Yes
Consistency																		
C1: internal consistency																		
Is there evidence that the mathematical logic of the model has been tested thoroughly before use?	Yes	N/A	N/A	Yes	Yes	Yes	N/A	N/A	Yes	Yes	N/A	Yes	N/A	N/A	N/A	N/A	N/A	N/A
C2: external consistency																		
Are the conclusions valid given the data presented?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Are any counterintuitive results from the model explained and justified?	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No
If the model has been calibrated against independent data, have any differences been explained and justified?	N/A	Yes	No	Yes	Yes	No	No	Yes	Yes	Yes	No	No	No	N/A	No	N/A	No	No
Have the results of the model been compared with those of previous models and any differences in results explained?	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No	No

C, consistency; D, data; N/A, not applicable; NR, not reported; S, structure.
Source: table II of Philips *et al.*[162]

Probabilities and costs of adverse events of diagnostic tests in included studies

TABLE 62 Probabilities and costs of adverse events of diagnostic tests in included studies

Study	Type (cost/utility/ probability)	Description	Value
Allen 2005[144]	Probability	Procedure complication rates for haemorrhage from colonoscopy	0.007
		Procedure complication rates for perforation from colonoscopy	0.004
		Mortality rate for colonoscopy	0.0001
		Procedure complication rates for perforation from FS	0.001
		Mortality rate for FS	0.00000006
		Procedure complication rates for perforation from ACBE	0.0001
		Mortality rate for ACBE	0.00002
	Cost	Costs of procedural complications of perforation	US$13,598
		Costs of procedural complications of haemorrhage	US$4561
Tappenden 2007[145]	Probability	Probability of perforation from colonoscopy (without polypectomy)	0.0008
		Probability of perforation from colonoscopy (with polypectomy)	0.0017
		Probability of death following perforation for colonoscopy	0.0582
		Probability of perforation (without polypectomy) from FS	0.000025
		Probability of perforation (with polypectomy) from FS	0.000025
		Probability of death following perforation for FS	0.0582
		Probability of bleeding following FS	0.000295
		Probability of bleeding following colonoscopy	0.00439
	Cost	Cost of treating bowel perforation (major surgery)	£5407.74
		Cost of admittance for bleeding	£250.21
Heitman 2010[148]	Probability	Risk of bleeding for diagnostic colonoscopy	0.0003
		Risk of bleeding for therapeutic colonoscopy	0.005
		Risk of perforation of diagnostic colonoscopy	0.0009
		Risk of perforation of therapeutic colonoscopy	0.0024
		Risk of perforation after FS	0.0002
		Risk of death after endoscopic perforation	0.049
	Cost	Cost of complication for perforation	CAN$31,223
		Cost of complication for bleeding	CAN$3194

TABLE 62 Probabilities and costs of adverse events of diagnostic tests in included studies (*continued*)

Study	Type (cost/utility/ probability)	Description	Value
Lee 2010[149]	Probability	Risk of bleeding from screening FOBT	0
		Risk of bleeding from screening FS	0.000295
		Risk of bleeding from screening colonoscopy	0.00439
		Risk of bleeding from screening CTC	0
		Risk of perforation from screening FOBT	0
		Risk of perforation from screening FS	0.000025
		Risk of perforation from screening colonoscopy	0.0008
		Risk of perforation from screening CTC	0.0006
		Risk of perforation from polypectomy FOBT	0.0017
		Risk of perforation from polypectomy FS	0.000025
		Risk of perforation from polypectomy colonoscopy	0.0017
		Risk of perforation from polypectomy CTC	0.0017
		Risk of mortality following perforation	0.00582
Knudsen 2012[150]	Probability	Risk of perforation from colonoscopy, by age group (years)	
		50–65	0.00036
		66–69	0.00036
		70–74	0.00042
		75–79	0.00052
		80–84	0.00064
		≥ 85	0.00087
		Risk of bleeding with transfusion from colonoscopy, by age group (years)	
		50–65	0.00089
		66–69	0.00089
		70–74	0.00103
		75–79	0.00127
		80–84	0.00156
		≥ 85	0.00214
		Risk of bleeding without transfusion from colonoscopy, by age group (years)	
		50–65	0.00245
		66–69	0.00245
		70–74	0.00284
		75–79	0.00351
		80–84	0.00430
		≥ 85	0.00589

continued

TABLE 62 Probabilities and costs of adverse events of diagnostic tests in included studies (*continued*)

Study	Type (cost/utility/ probability)	Description	Value
		Risk of other gastrointestinal events from colonoscopy, by age group (years)	
		50–65	0.00320
		66–69	0.00320
		70–74	0.00400
		75–79	0.00540
		80–84	0.00730
		≥ 85	0.00880
		Risk of perforation from CTC, by age group (years)	
		50–65	0.00005
		66–69	0.00005
		70–74	0.00005
		75–79	0.00005
		80–84	0.00005
		≥ 85	0.00005
	Cost	Cost of perforation from colonoscopy	US$15,985
		Cost of bleeding with transfusion from colonoscopy	US$7784
		Cost of bleeding without transfusion from colonoscopy	US$1775
		Cost of other gastrointestinal events from colonoscopy	US$1195
		Cost of perforation from CTC	US$15,985
Sharp 2012[151]	Probability	Probability of perforation from FS with polypectomy	0.00002
		Probability of perforation from FS without polypectomy	0.00002
		Probability of death after perforation FS	0.06452
		Probability of (major) bleeding following FS	0.00029
		Probability of perforation from colonoscopy with polypectomy	0.00216
		Probability of perforation from colonoscopy without polypectomy	0.00107
		Probability of death after following perforation	0.05195
		Probability of (major) bleeding following colonoscopy	0.00379
	Cost	Cost of treating bowel perforation	€10,200
		Cost of admittance for bleeding	€3079
Whyte 2012[152]	Probability	Perforation rate from colonoscopy with polypectomy	0.003
		Perforation rate from colonoscopy without polypectomy	0
		Probability of death after perforation colonoscopy	0.052

TABLE 62 Probabilities and costs of adverse events of diagnostic tests in included studies (*continued*)

Study	Type (cost/utility/ probability)	Description	Value
		Perforation rate from FS with polypectomy	0.001
		Perforation rate from FS without polypectomy	0
		Probability of death after perforation FS	0.065
		Probability of hospitalisation for bleeding after FS	0.0003
		Probability of hospitalisation for bleeding after colonoscopy	0.003
	Cost	Cost of treating bowel perforation (major surgery)	£2164
		Cost of admittance for bleeding (overnight stay on medical ward)	£278
Goede 2013[153]	Probability	Fatal complications rate after colonoscopy	0.0001
	Cost	Costs complications after colonoscopy	€1250
Gomes 2013[154]	Probability	Rate of complications from colonoscopy bleeding	0.0044
		Rate of complications from colonoscopy perforation	0.002
		Rate of polypectomy bleeding rate	0.02
		Rate of polypectomy perforation	0.0038
		FS bleeding rate	0.0003
		Risk of death after perforation	0.0582
	Cost	Costs of admittance for bleeding	£295
		Cost of treatment of bowel perforation	£5822
Tappenden 2013[155]	Probability	Probability of perforation from colonoscopy	0.0013
		Probability surgery for perforation	0.66
		Probability FS perforation	0.2×10^{-4}
		Probability DCBE perforation	0.4×10^{-4}
		Probability CTC perforation	0.0005
		Probability patient dies during perforation surgery	0.06
		Probability patient dies during polyp surgery	0.04
		Probability patient undergoes stenting	0.20
		Probability stenting successful	0.72
		Probability stent-related mortality	0.0025
	Cost	Emergency surgery postoperative mortality rate	£294.77
		Cost emergency surgery	£7079.93
		Cost emergency surgery for perforation	£5088.91
		Cost conservative management of perforation	£881.99
		Cost stent insertion	£1012.73

continued

TABLE 62 Probabilities and costs of adverse events of diagnostic tests in included studies (*continued*)

Study	Type (cost/utility/ probability)	Description	Value
Whyte 2014[156]	Probability	Probability of perforation rate of colonoscopy (with polypectomy)	0.003
		Probability of perforation rate of colonoscopy (without polypectomy)	0
		Probability of death following perforation from colonoscopy	0.052
		Probability of hospitalisation for bleeding from colonoscopy	0.003
	Cost	Cost of treating bowel perforation (major surgery)	£5089
		Cost of admittance for bleeding (overnight stay on medical ward)	£278
Cantor 2015[157]	Not included	Not included	
Pil 2016[158]	Not included	Not included	
Wong 2016[159]	Probability	Bleeding rate	0.002
		Perforation rate	0.0002
		Morality due to perforation	0.1
	Cost	Cost of bleeding	US$3320
		Cost of perforation	US$10,790
Coldman 2017[160]	Not included	Not included	
Westwood 2017[161]	Probability	Bleeding rate after colonoscopy	0.0026
		Perforation rate after colonoscopy	0.0005
		Death rate after colonoscopy	0.000029
	Cost	Cost of bleeding	£603
		Cost of perforation	£3228
		Cost of death	£0

ACBE, air-contrast barium enema; DCBE, double-contrast barium enema; FS, flexible sigmoidoscopy.

Decision model data and assumptions

Clinical effectiveness parameters

Distribution of disease stages at presentation

As, by definition, no data exist on the distribution of symptomatic patients by disease stage at first presentation to the GP, we replicated the disease history model by Tappenden *et al.*[145] and generated the distribution of disease stages at presentation by age from that model. The model by Tappenden *et al.*[143] was calibrated to data on CRC prevalence and incidence (see *Chapter 6*), and we verified that predictions from our replication of that model matched the national mortality data on CRC incidence in England.[398]

Diagnostic accuracy of the faecal immunochemical test

A systematic review of diagnostic accuracy for FIT found no diagnostic accuracy studies for FIT at thresholds of 10 µg Hb/g faeces and 20 µg Hb/g faeces evaluated in the low-risk population of interest here, which was defined by the NICE Guideline Number 12[6] as that with a prevalence of 0.1–3%.[161] Furthermore, only one study of those identified was conducted in a primary care setting, but, in this

study, the FIT was evaluated in patients who were about to be referred to secondary care. The meta-analysed values from four studies for FITs of 10 µg Hb/g faeces with the OC-Sensor (Palex Medical S.A., Sant Cugat del Vallès, Spain) were 92.1% (86.9% to 95.3%) for sensitivity and 85.8% (78.3% to 91.0%) for specificity;[161] sensitivity and specificity of 100% and 76.6%, respectively, for HM-JACKarc (Hitachi Chemical Co., Ltd, Tokyo, Japan) from one study were also reported, but are not considered here as they were based on only 11 CRC cases. As these estimates may overestimate the sensitivity of the test, we use these values for the sensitivity analysis. For our base-case analysis, we chose to use instead the values reported by Murphy *et al.*[175] for a screening programme for which a FIT of 20 µg Hb/g faeces, the threshold with the highest sensitivity of those investigated, was reported to have sensitivity of 52.6% and specificity of 98.8% for those aged 50–69 years, and specificity of 96.3% for those aged ≥ 70 years.

Excess mortality of symptomatic undiagnosed patients

Patients with false-negative results are exposed to delays in diagnosis and treatment, and we accounted for the excess risks of death from lack of treatment experienced by an as yet undiagnosed CRC patient. The only source of evidence that we identified was a study of newly diagnosed CRC patients who refused treatment in Hong Kong and whose survival outcomes were compared with patients who accepted treatment for CRC.[177] The study subjects included patients with CRC at disease stages I–IV, in whom a hazard ratio death of 2.66 (95% CI 2.49 to 2.84) was reported for refusing treatment. The study did not report a treatment effect by disease stage. Therefore, we used this estimate for all disease stages except for Dukes' stage A, for which we conservatively assume that lack of treatment has no immediate effect on survival, in line with previous models.[145]

Relative mortality of symptomatic undiagnosed patients

We estimate the CRC-related death probabilities from the relative survival curves by Dukes' stage depicted in *Figure 9*. We digitised the curves and fitted exponential survival models to the extracted data (R^2 values of 0.98, 0.99, 0.99 and 0.97 for relative survival curves of Dukes' stages A, B, C and D, respectively). The resulting estimates are presented in *Tables 6* and *7*. Admittedly, these estimates, which were obtained from data of patients diagnosed between 1996 and 2002, are outdated, as 5-year survival in CRC patients has increased since the time the data were collected. However, this is the most recent available data by disease stage, and further sensitivity analyses revealed that this parameter had limited influence on the study results.

It must be noted that using the same data Whyte *et al.*[152] have reported net survival curves by disease stage at diagnosis and age. Given the time constraints we faced and the limited influence that these age-specific CRC survival data would have on the results, we did not incorporate these data into the model. We note that, by using relative survival estimates instead of net survival ones, we implicitly assume that there is no heterogeneity in survival outcomes, which is a strong assumption, but one with no consequence to our results.

Costs

Faecal immunochemical test

Our unit cost estimate for a FIT was obtained from a screening programme evaluation.[161] The figure used for the model excludes the costs of research. This is lower than the £4.53 and £6.03 figures for OC-Sensor and HM-JACKarc, respectively, used by a previous study.[161] We explore the effect of these alternative cost estimates in sensitivity analyses.

Colonoscopy

We obtained the costs of colonoscopy from the data reported by Whyte and Harnan.[156] The original figure of £563 was obtained from NHS reference costs[399] and data from a single screening centre.[156] The figure reflated to 2018 prices used in the model for this parameter is presented in *Tables 6* and *7*. A previous study adopted a much lower cost of £372 from an average of outpatient diagnostic colonoscopy with and without biopsy in NHS reference costs of 2014–15.[161,400] We retained the higher cost for our base-case analysis and used the lower cost for the sensitivity analysis.

We use the same source to account for the costs of adverse events with colonoscopy, namely perforation and bleeding.[156] The NHS reference costs reported by the source, £5089 and £278 for perforation and bleeding, respectively, are different from those used in the study by Westwood *et al.*[161] of £3228 and £603, respectively, which were derived from length of hospital stay estimates of those events reported in a 2012 study (1.7 and 9.1 days, respectively[161,401]). We use the former set of figures for the base-case analysis, and the latter for the sensitivity analysis.

Treatment of colorectal cancer

Lifetime costs of treatment by Dukes' disease stage and age were obtained from a previous study.[152] This source derived its estimates from simulations of the disease history model by Tappenden *et al.*[145] The estimates used in the model are those figures reflated to 2018 prices, and presented in *Tables 6* and *7*. Although no better source of estimates was identified, the value of these parameters was found to have negligible influence on the study results.

Utilities

We adopted utility values by CRC disease stage used in the model by Tappenden *et al.*,[145] which obtained them from the study by Ness *et al.*[184] These values were used by the existing study in the field.[161] These values were derived from individual interviews of 81 patients who had previously undergone removal of colorectal adenoma; during the interviews, participants were presented with stage-dependent outcome states and asked to assess their relative value using the standard gamble technique. In the absence of utility data for treated patients, we assumed that these values are applicable to both untreated and treated patients. In sensitivity analyses, we found that the choice of values for these parameters did not affect the result in a significant way.

In addition, utility values for both patients with CRC and healthy individuals are applied background utilities from the model by Ara and Brazier,[402] which accounts for the non-linear decline of utility with age, and differentiates between age and sex.

We found no available estimate on the negative impact on health-related quality of life of colonoscopy or its adverse events. Therefore, we limited our analysis to explore the effect of a disutility of colonoscopy of 0.0075 in the month when referral takes place, that is the equivalent of a loss of life in full health of 5 hours.

Assumptions in Westwood *et al.*'s[161] model versus assumptions in our model

Table 63 presents a comparison of the assumptions used in the previous diagnostic model[161] relevant to our analysis and our own model. Our modelled cycle length is 28 days, as opposed to Westwood *et al.*'s[161] yearly length, which is too long to capture the small differences in the referral interval between the diagnostic strategies that we investigated. We modelled the probability of delayed referral and diagnosis in patients with a false-negative outcome after the initial tests in primary care, as opposed to the previous model's[161] arbitrary assumption of a 6-month delay to diagnosis. In contrast to Westwood *et al.*'s[161] model, in our model an account is made for excess mortality during the time spent in a false-negative state.

In terms of resource use, our cost of colonoscopy (£596 in 2014/15 prices), based on Whyte and Harnan,[156] is higher than that used by Westwood *et al.*[161] (£372), despite the fact that both are derived from NHS reference costs. However, the difference is reduced to £89 because Westwood *et al.*[161] include an additional cost of £135 for a follow-up appointment with a gastroenterologist to discuss the colonoscopy results, which is not explicitly considered by Whyte and Harnan[156] (who provide no further detail). Unlike the Westwood *et al.*[161] model, investigations for determining disease stage in patients diagnosed with CRC are not considered in our analysis, but sensitivity analyses indicate that these have a negligible effect on the results.

TABLE 63 Assumptions in Westwood *et al.*'s[161] model vs. assumptions in our model

Assumption number	Westwood *et al.*[161] model	Our model
1	A lifetime horizon with a 1-year cycle length captures the probability of progression for treated and untreated patients	Lifetime horizon with 4-weekly cycles is required to measure outcomes with sufficient accuracy
2	Any differences in costs between the tests in patients without CRC were assumed to occur only in the first year	Any differences in costs between the tests in patients without CRC were assumed to occur only in the first month
3	Any differences in life expectancy between intervention and comparator for patients without CRC are due only to the difference in mortality due to colonoscopy/CTC	Same assumption
4	Any differences in costs between intervention and comparator for patients without CRC are only due to difference in cost of gFOBTs and colonoscopy/CTC	Any differences in costs between intervention and comparator for patients without CRC are only due to difference in cost of FIT and colonoscopy
5	Testing has no long-term (after 1 year) effect on costs or QALYs in disease-negative people. Thus, in patients without CRC, FOBTs would not significantly delay diagnosis of the underlying cause of presenting symptoms, and hence would not incur any extra cost or effect on mortality	Same assumption but applies after 1 month
Diagnostic model		
6	A time frame of 1 year was assumed for the diagnostic model	Integrated within Markov model
7	A positive FIT/gFOBT result in referral to colonoscopy	A positive (above the risk threshold) tool assessment results in referral to colonoscopy, as does a positive FIT result after a negative tool result in the intervention or, in the control arm, in all patients
8	A negative FIT or gFOBT result in a watchful waiting strategy, in which a colonoscopy/CTC will be performed when symptoms persist. A repeat FIT/gFOBT might also be performed, but referral to colonoscopy/CTC is not modelled as being contingent on the results of the repeat test	A negative FIT, in the control, or both tool and FIT, in the intervention, results in a watchful waiting strategy, in which repeat testing will be performed when symptoms persist (which is assumed to occur in all CRC cases and none of the non-CRC patients), and referral to colonoscopy is modelled as being contingent on the results of the repeat testing
9	The sensitivity and specificity of colonoscopy for detection of CRC is 100%	Same assumption
10	The symptoms of all those patients with CRC who receive a false-negative result will persist such that they will all receive a colonoscopy and thus be diagnosed (within 1 year) should they survive (NG151[6]/expert opinion[403])	Same assumption, but < 1% of patients will receive colonoscopy beyond the 1 year – as determined by the probability of a positive result after repeated monthly trials of repeat visit and testing according to sensitivity of the test. This corresponds with data available to authors from the electronic primary care record
11	Patients who had a false-negative gFOBT or FIT result, and whose symptoms persisted, have an increased probability of progressing to a worse cancer state because of the delay in diagnosis	Same, all false-negative patients have persistent symptoms
12	Probability of delayed diagnosis of CRC was assumed to be the probability of progression within Dukes' stages at 6 months	The probability of time to diagnosis was determined by the sensitivity of the overall strategy, with the probability of diagnosis at cycle t being given by the Bernoulli formula for 'success' out of t trials (cycles), that is $[(1 - \text{sensitivity})^{t-1}] \times \text{sensitivity}$, which, when combined with the rate of disease progression over Dukes' disease stages, determined the disease stage at diagnosis

continued

TABLE 63 Assumptions in Westwood *et al.*'s[161] model vs. assumptions in our model (*continued*)

Assumption number	Westwood *et al.*[161] model	Our model
14	Only those patients with a negative test result, and whose symptoms do not persist, do not receive a colonoscopy/CTC	Patients with a negative test result who have CRC have persistent symptoms and return every month for testing and get referred to receive colonoscopy with a probability determined by the sensitivity of the strategy (see 13). All negative-testing patients without CRC are assumed to clear the symptoms after the initial test result and not to return to the GP
15	CT of the chest, abdomen and pelvis is performed for all of the patients testing positive for CRC after colonoscopy or CTC, to estimate the stage (Dukes' A–D) of the disease (NG12[6])	This cost element is not considered, but has a negligible effect on the results
16	For the base-case scenario, we considered a threshold of 10 µg Hb/g faeces, or equivalent, for the detection of CRC using a single faecal sample. Other options were explored in sensitivity analyses	For the base-case scenario, we considered a threshold of 20 µg Hb/g faeces, as the evidence on 10 µg Hb/g faeces was derived from a population different from that of interest
17	The prevalence of CRC in the base-case population was 1.5%, as in NG12[6]	The prevalence of CRC in the base-case population was 1.5%
CRC Markov model		
18	After the initial distribution of patients in the CRC model is determined, patients may stay in their current health state, progress to the health state representing the next worsening in the condition or die (from CRC or another cause) (NG12[6])	Same specification
19	Costs associated with the health states of the CRC Markov model were estimated as lifetime costs (i.e. one-off cost)	Same specification
Healthy population Markov model		
20	Patients entering this model can either die of all of the causes or stay in the 'alive' health state	Same specification
21	For the base-case scenario, patients aged ≥ 40 years (NG12[6])	For the base-case analysis, patients aged 70 years
22	The adverse events included in the diagnostic model are bowel bleeding, perforation and death (literature)	Same specification
Adverse events		
23	Reduction of quality of life due to adverse events is assumed to be negligible within a lifetime (assumption)	Same assumption
24	No costs of patients who die owing to adverse events of colonoscopy	Same assumption
25	Adverse events due to CTC were assumed to be the same as those of colonoscopy	CTC was not included in the model
Test costs		
26	Costs of laboratory staff to analyse the test were assumed to be the same for FITs and gFOBTs	Irrelevant; FOBT was not considered as is not used in practice
27	Training costs and the costs of the laboratory staff for analysing the test results were not included in the total costs because it was assumed that these are the same for FITs and gFOBTs	Laboratory staff costs for analysis were included in the cost of a FIT; training costs were not included, to correspond with a measure of marginal costs
28	Costs of the material needed to analyse a sample include costs of the reagents, buffer, reaction cells and analyser cups	Same cost elements were included

TABLE 63 Assumptions in Westwood *et al.*'s[161] model vs. assumptions in our model (*continued*)

Assumption number	Westwood *et al.*[161] model	Our model
29	Costs of colonoscopy/CTC, adverse event costs and CT costs were included in the diagnostic model (NG12[6])	Same costs except CTC were included in the model
30	We assumed that the cost of colonoscopy/CTC includes the costs of a follow-up appointment with a gastroenterologist (assumption/expert opinion)	The cost of a follow-up appointment with gastroenterologist is implicit in our chosen unit cost of colonoscopy (£563 in 2011 prices, £596 after reflation to 2014–15 prices), which was taken from Whyte and Harnan[156] and compares with Westwood *et al.*'s[161] £372 (in 2014–15 prices); the difference (£224) is £89 higher than the £135 unit cost of a gastroenterology outpatient appointment used by Westwood *et al.*[161] Whyte and Harnan's[156] reported source for the cost of colonoscopy is the NHS reference costs (no further details provided), as is Westwood *et al.*'s.[161]
31	For test-negative patients whose symptoms persist, an additional GP appointment cost was considered	Negative-testing patients whose symptoms persist are all false negatives and keep returning each month to the GP until they test positive or die without referral and diagnosis
32	Indirect costs parameters were not included in the model, given the perspective of the NHS	Same specification
33	Zero excess mortality during delayed diagnosis in CRC patients	Positive excess mortality during delayed diagnosis in CRC patients

CT, computerised tomography; NG NICE Guideline.

Sensitivity analyses

TABLE 64 Mean time to referral of CRC patients by strategy in the model

	QCancer		Comparator[a]		RAT	
Referral after	Predicted frequency[b,c]	Mean time to referral (days)	Predicted frequency[c,d]	Mean time to referral (days)	Predicted frequency[b,c]	Mean time to referral (days)
1 visit	0.49	28	N/A	28	0.55	28
2 visits	0.24	28	0.61	28	0.19	28
3 visits	0.13	56	N/A[e]	N/A	0.14	56
4 visits	0.07	56	0.24	56	0.05	56
≥ 5 visits	0.08	95	0.15	102	0.07	94
Weighted average	1	39	1	46	1	38

N/A, not applicable.

a The modelled comparator strategy assumes that, every time the patient returns, two GP visits are required for a referral decision to be made: one for the return visit and a second, follow-up visit for the review and discussion of the results of primary care investigations ordered at the return visit.

b Positive test results with the tool result in referral after the first visit, whereas a positive test result with the FIT after a negative result with the tool takes two visits.

c Derived from the sensitivity values for the component tests times the 80% rate of compliance of referral practice with the test results, see *Table 19*.

d Positive test result with FIT alone results in referral after two visits: an initial visit and a follow-up visit to discuss results.

e Values based on implicit sensitivity of 0.76 in Lyratzopoulos *et al.*[173]

TABLE 65 Disease state at diagnosis of CRC patients by strategy in the model

	QCancer		Comparator[a]		RAT	
Referral after	Predicted frequency	Dukes' A/B (%)	Predicted frequency	Dukes' A/B (%)	Predicted frequency	Dukes' A/B (%)
1/2 visits	0.73	54	0.61	54	0.74	54
3/4 visits	0.20	51	0.24	51	0.19	51
5/6 visits	0.05	48	0.09	48	0.05	48
7/8 visits	0.01	45	0.04	45	0.01	45
9/10 visits	0.00	42	0.01	42	0.00	42
> 10 visits	0.00	40	0.01	≤ 40	0.00	40
Weighted average	1	53	1	52	1	53

a Other parameters were set at base-values presented in *Table 19* and it is at the end of the title of the table.

Probabilistic sensitivity analyses

TABLE 66 Parameters varied in probabilistic sensitivity analysis, the base case was for a 70-year-old individual and their distributions[a]

Parameter	Value	Distribution	Source
Clinical			
Prevalence of CRC	0.015	Beta(1838,121683)	Hippsley-Cox 2012[85]
Sensitivity of QCancer	0.610	Beta(1198,765)	Hippsley-Cox 2012[85]
Specificity of QCancer	0.910	Beta(1111261,109968)	
Sensitivity of FIT	0.526	Beta(73,65)	FIT of 20 µg Hb/g faeces; Murphy 2017[175]
Specificity of FIT: 50–69 years old	0.988	Beta(10148,123)	
Specificity FIT: ≥ 70 years old	0.963	Beta(9891,380)	
Compliance with colonoscopy referral	0.80	Beta(80,20)	Assumption
Transition probabilities		4-weekly	Whyte 2014[156]
Dukes' A to Dukes' B	0.0651	Beta(651,9349)	Assumption. Tappenden 2007[145] reported model calibration to CRC incidence and stage at diagnosis
Dukes' B to Dukes' C	0.0787	Beta(787,9213)	
Dukes' C to Dukes' D	0.1427	Beta(1427,8573)	
Costs			
Cost of colonoscopy	£615	Uniform(555,676)	Assumption, ± 10% following Whyte 2011,[404] appendix 4

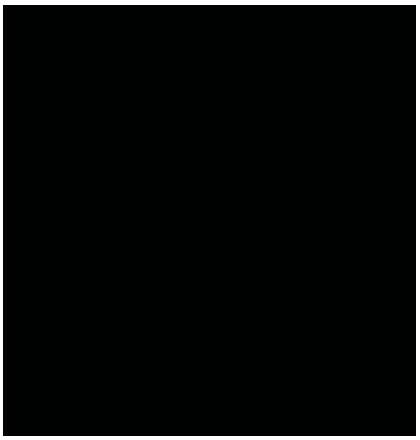
a Other parameters were set at base-case values presented in *Table 19*.

Appendix 6 General practitioner survey

Cover letter

To:

Dr Sarah Price



18 March 2020

Re: Use of cancer diagnostic tools to aid decision-making in primary care

Dear

My name is Dr Sarah Price and I am a research fellow at University of Exeter Medical School. I am writing to ask for your help with our study of the use of cancer diagnostic tools in primary care, which is funded by National Institute of Health Research from the Health Technology Assessment Programme (Grant 16/12/04). One of the elements of the study is conducting a survey of UK GPs and asking them about their access and use of cancer diagnostic tools.

Please distribute the enclosed questionnaires to each of the GPs in your practice. The GPs will return their completed surveys to you, and I'd be grateful if you would package them up in the enclosed business reply envelope and return them to Binley's by DATE.

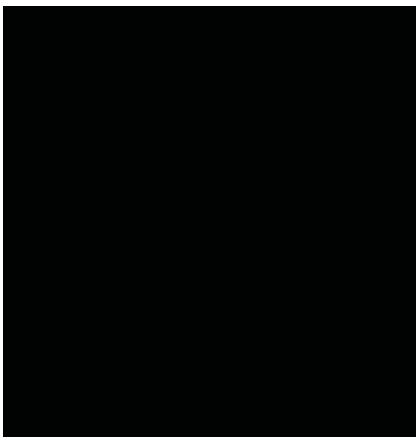
Yours sincerely,

Dr Sarah Price

Participant information

To:

Dr Sarah Price



18 March 2020

Use of cancer diagnostic tools to aid decision-making in primary care

Dear Dr

We are writing to request your participation in our study of use of cancer diagnostic tools in primary care. It involves completing a short, anonymised postal questionnaire. Please read this information sheet carefully before deciding whether or not to participate. If you decide to participate we thank you; if not, we thank you for considering our request.

What is the aim of the project?

The main aim of this project is to obtain an estimate of how many GPs in the UK use cancer diagnostic support tools. We also want to gain an understanding of how the tools are used. Finally, we want to compare cancer referral rates of GP practices that report using the tools with referral rates of GP practices that do not. The project is being undertaken by researchers from the University of Exeter Medical School and is funded by NIHR Health Technology Assessment Grant 16/12/04.

Description of participants required

The participants are a representative sample of GPs in the UK. The sample of GPs was provided by Binley's (www.binleys.com), a commercial company that has an up-to-date listing of GPs in the UK and supports the Royal College of General Practitioners in maintaining its database.

What will participants be asked to do?

Should you agree to take part in this study, you will be asked to:

Read this cover letter, explaining the purpose and nature of the questionnaire

Answer 8 questions relating to cancer decision support tools and 2 questions asking you to confirm your number of years in practice and the number of sessions per week worked

Give the completed questionnaire to your practice manager, who will return it to us using a prepaid envelope

Time commitment

The questionnaire is anticipated to take 5–10 minutes to complete.

Payment/reward to volunteers/interviewees

There is no payment to GPs for taking part in this research. We will make a £7.50 donation to a cancer charity for each completed questionnaire, capped at 400 questionnaires.

Can participants change their mind and withdraw from the Project?

You may withdraw from participation in the project at any time without any disadvantage to yourself of any kind. Consent is assumed by your completion and return of the survey.

What data or information will be collected and what use will be made of it?

The responses to the questionnaires will be collected, along with details of the GP software used at your practice (as supplied by Binley's). The data is being collected with the purpose of enabling us to meet the aims of our study as outlined above. The data will be anonymised and collated by Binley's, and then analysed by the principal researcher (Sarah Price). The anonymised analysis will be accessible to the following researchers at University of Exeter:

- Sarah Price (principal researcher)
- Antonieta Medina-Lara (lead investigator, Senior Lecturer in Health Economics)
- Anne Spencer (co-investigator, Associate Professor of Health Economics)

The data collected will be securely stored so that only those mentioned above will be able to gain access to it. The aim is to publish the results of this project under open access, and so freely available to participants in this study. If you wish to receive an electronic copy of the results, please confirm

this by emailing Sarah Price (at S.J.Price@exeter.ac.uk). Any data included will not be individually identifiable.

Why me?

We aim to obtain a representative sample of GPs in the UK for our study, so that we can make inferences about our results to the whole country. You were selected at random from all practising GPs in the UK.

What if participants have any questions?

If you have any questions about our project, either now or in the future, please feel free to contact either:

Dr Sarah Price	Dr Antonieta Medina-Lara
Research Fellow	Senior Lecturer

[redacted]

Complaints

If you have any complaints about the way in which this study has been carried out please contact the Chair of the University of Exeter Medical School Research Ethics Committee:-

Ruth Garside, PhD

Chair of the UEMS Research Ethics Committee

[redacted]

Funding acknowledgement

This research was funded by the NIHR Health Technology Assessment grant 16/12/04.

With grateful thanks,

[redacted]

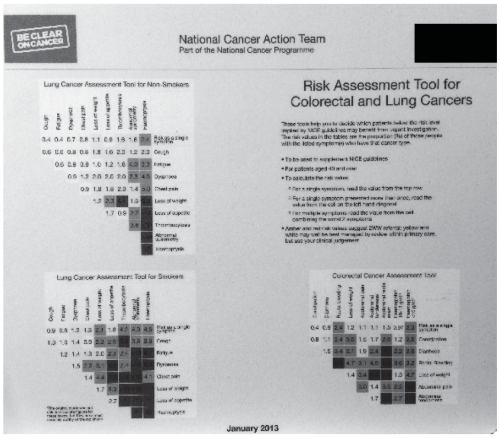
Aids to cancer diagnosis in primary care survey

Please give your completed survey to your practice manager by **17 August 2017** who will return it to us using a pre-paid envelope. **Consent to participate is assumed by return of your completed survey.**

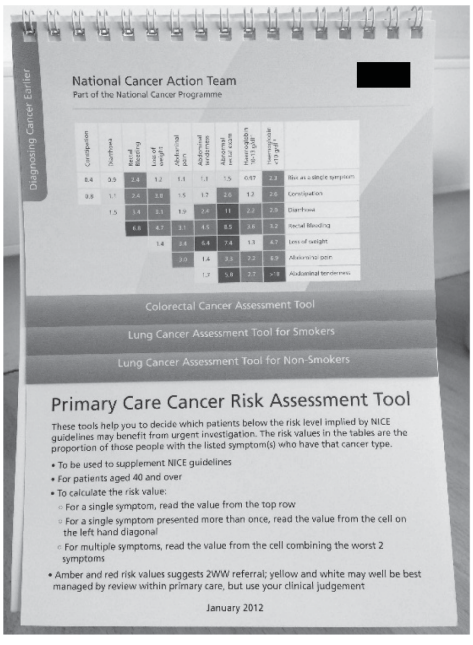
Cancer risk assessment tools – in desktop and electronic forms – are available to help GPs estimate the risk of cancer in symptomatic patients.

1. **Two examples of desktop risk assessment tools are pictured below. Please tick those that you have in your consulting room**

☐ Mousemat



☐ Flipchart



☐ Other (please describe)

☐ Tick here and **move on to Q4** if you have no desktop risk assessment tools.

2. A patient presents with symptoms that might be due to cancer. How likely are you to consult a desktop risk assessment tool?

☐ Very likely ☐ Likely ☐ Unlikely ☐ Very unlikely

3. Which functions of the desktop tools do you find helpful? Select as many as appropriate

☐ Assessing cancer risk in patients with non-specific symptoms

☐ Assessing cancer risk in patients with multiple symptoms

☐ Discussing cancer risk with a patient

☐ Discussing investigation with symptomatic patients

☐ Prompting referrals that I would otherwise not have made

☐ Increasing the certainty of my clinical decision-making

☐ Increasing my awareness of cancer as a possible diagnosis

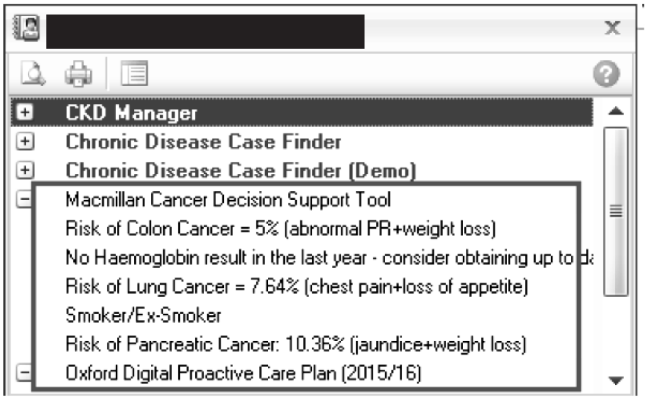
☐ Increasing my awareness of cancer symptoms

☐ Reassuring anxious patients

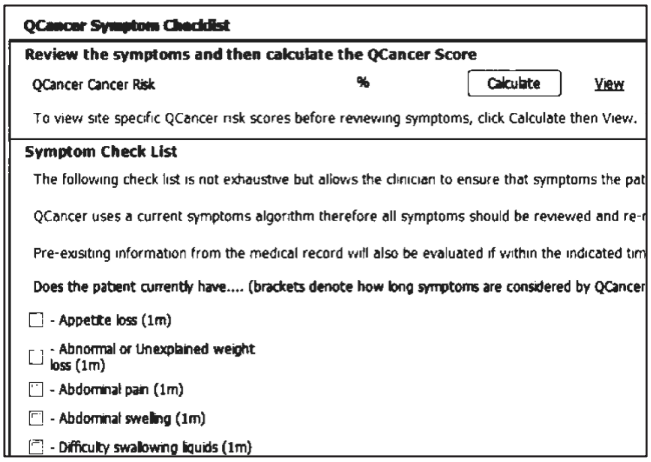
☐ None of these

4. **Macmillan have produced electronic cancer decision support (eCDS) tools to help GPs estimate cancer risk in symptomatic patients.**

In Vision (RAT algorithm)



In EMIS (QCancer®)



Please select ONE option below that best reflects eCDS availability at your practice, or tick here ☐ and move on to Q9 if you are unaware of eCDS.

☐ eCDS is downloaded/activated for my IT system

☐ eCDS is available for my IT system, but my practice has not downloaded/activated it

☐ eCDS is available for my IT system, and my practice has plans to download/activate it in future

☐ To my knowledge, eCDS is not available for my IT system

☐ eCDS is not available for my IT system but I would like to have it

5. **eCDS utilisation – please select one option**

☐ I have received eCDS training and have integrated the tool into my practice

☐ I have received eCDS training but have not integrated the tool into my practice

☐ I have not yet received eCDS training, but have integrated the tool into my practice

☐ I have not yet received eCDS training, but it is planned

☐ I have not yet received eCDS training, and I'm not aware that any is planned

6. Please rank the 3 main functions of eCDS numerically in order of usefulness (1=most useful; 3=least useful)

☐ 'Alert/Prompt' function: cancer risk scores appear automatically on opening the patient's record

☐ 'Cancer risk assessment' function: GPs can request a patient's cancer risk score using the symptom checker

☐ 'Searches/report' function: for routinely searching records and producing reports for safety-netting purposes

☐ Not applicable: I have not integrated eCDS into my practice

7. Which other eCDS functions do you find helpful? Select as many as appropriate

☐ Assessing cancer risk in patients with non-specific symptoms

☐ Assessing cancer risk in patients with multiple symptoms

☐ Discussing cancer risk with a patient

☐ Discussing investigation with symptomatic patients

☐ Prompting referrals that I would otherwise not have made

☐ Increasing the certainty of my clinical decision-making

☐ Increasing my awareness of cancer as a possible diagnosis

☐ Increasing my awareness of cancer symptoms

☐ Reassuring anxious patients

☐ None of these

8. A patient presents with symptoms that may be caused by cancer. How likely are you to access eCDS? Please select one option

☐ Very likely ☐ Likely ☐ Unlikely ☐ Very unlikely

9. How many years have you been practising as a GP?

☐ <1 ☐ 1–5 ☐ 6–10 ☐ 11–20 ☐ 21–30 ☐ >30

10. How many consulting sessions per week do you work at this practice?

☐ 1–2 ☐ 3–4 ☐ 5–6 ☐ 7–8 ☐ 9–10 ☐ >10

Thank you for taking the time to fill out our survey

Funding acknowledgement: This research was funded by the NIHR Health Technology Assessment grant 16/12/04.

Other types of cancer decision support tools named by general practitioners in the free text as being available to them

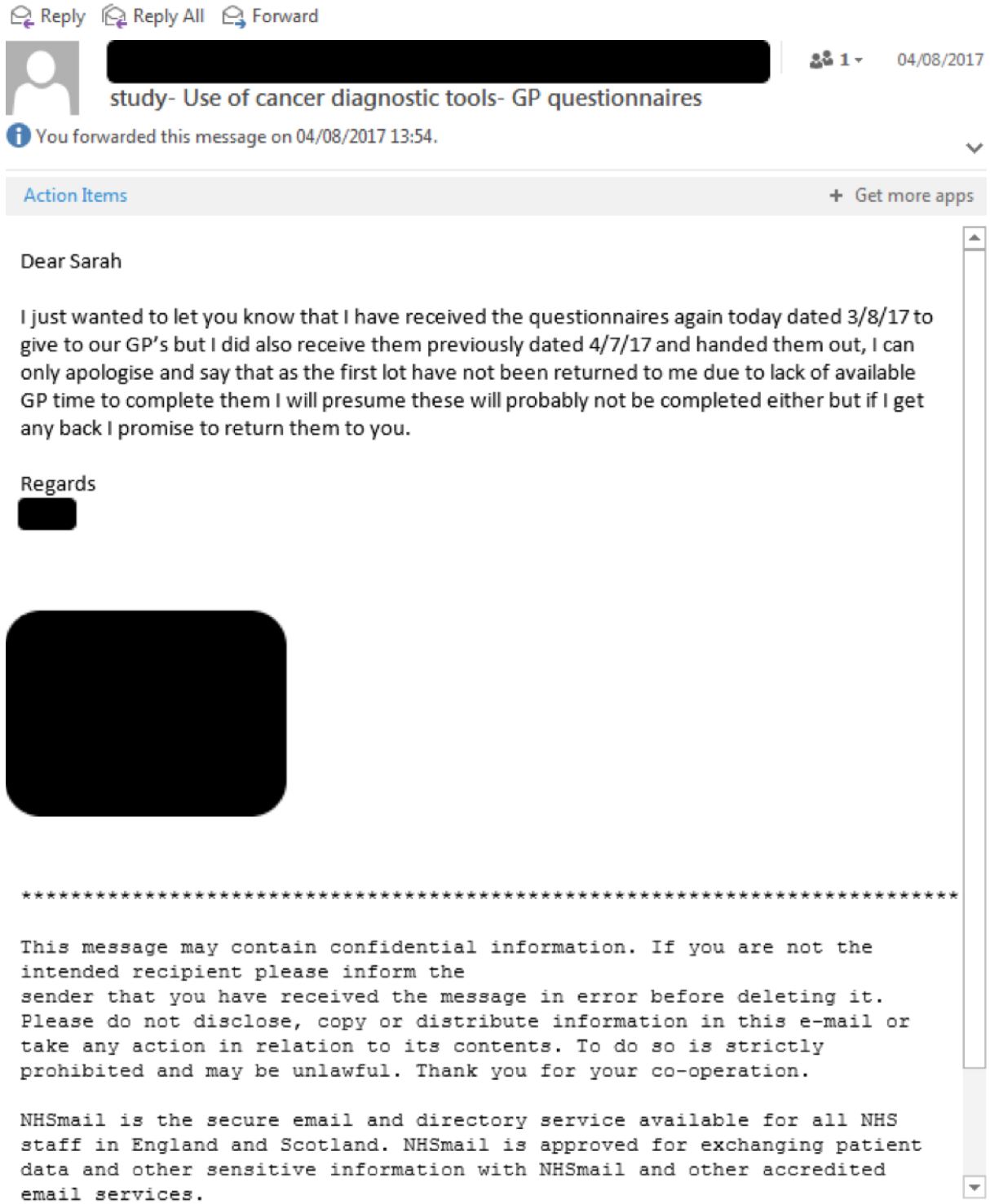
Cancer guideline summaries

- BMJ cancer guidelines chart
- BMJ flow charts on computer desktop
- BMJ Informatica
- Booklet from a GP 'Update' course
- Fast-track referral guidance/NICE guidelines
- GP update summary of cancer guidelines NICE 2015
- Macmillan Rapid Referral guidelines
- NICE guidelines 2WW form
- NICE/BMJ cancer assessment/referral guidelines poster
- NICE: suspected cancer recognition and referral
- Pathfinder (Northants)
- Scottish Referral Guidelines for suspected cancer (quick reference guide)
- Secure Grampian Cancer Network chart (NHS Grampian)

Comments received

- GP unable to complete at present.
- GP unable to complete at present.
- GP unable to complete at present.
- GP unable to complete at present.
- Registrar – unable to complete at present.
- GPs do not use these tools.
- GPs do not use these tools.
- GPs do not use these tools.
- GPs do not use these tools.
- GPs do not use these tools.
- GPs do not use these tools.
- Registrar + 3 Year of military primary care experience.
- Unfortunately no time.
- Incomplete as GP has left.
- Q1 comment – Have Macmillan guidance.
- Tool itself was excellent. Did have Macmillan Tool kit for few months but abandoned – IT (local) support inadequate.
- The surgery have advised that they have not heard of the cancer tools so will not complete the survey.
- Never heard of QCancer I was unaware of it but having checked the computer it is there and available.
- Likely if I am told that it picks up more cancer or reduces referrals in trials using eCDS. Also likely if, when using, the tool, I can quote on the referral that the eCDS recommends referral and this would be actioned by secondary care and not ignored.
- The practice have advised that they use SystmOne so do not have acess to the electronic cancer decision support tools. Because of this they will not be returning the survey.
- I haven't been using one regularly but would be interested in using a mousemat tool.
- Sorry GPs do not wish to participate.
- Declined.
- Q7 – I use it off the internet QCancer.
- Q2 – would if I had one – mousemat would be useful.
- Q4 – I tried to use when I was a GP in Wirral on EMIS Web but not successful & Q6 – I have seen in action at Macmillan conference.
- Q1 – Did have mousemat but don't any longer.
- Not completed – too time-consuming.
- Not completed – too time-consuming.
- Not completed – too time-consuming.
- Not completed – too time-consuming.
- I raised this as an option & past partners felt concerned that not acting on a certain risk would make them more likely to be sued. It never went any further.
- Flipchart – had one but too much paper – not practical for daily use, too 'fussy'.

Copy of email received


Reply Reply All Forward

[redacted] 1 04/08/2017

study- Use of cancer diagnostic tools- GP questionnaires

You forwarded this message on 04/08/2017 13:54.

Action Items + Get more apps

Dear Sarah

I just wanted to let you know that I have received the questionnaires again today dated 3/8/17 to give to our GP's but I did also receive them previously dated 4/7/17 and handed them out, I can only apologise and say that as the first lot have not been returned to me due to lack of available GP time to complete them I will presume these will probably not be completed either but if I get any back I promise to return them to you.

Regards

[redacted]

[redacted]

****

This message may contain confidential information. If you are not the intended recipient please inform the
sender that you have received the message in error before deleting it.
Please do not disclose, copy or distribute information in this e-mail or take any action in relation to its contents. To do so is strictly prohibited and may be unlawful. Thank you for your co-operation.

NHSmail is the secure email and directory service available for all NHS staff in England and Scotland. NHSmail is approved for exchanging patient data and other sensitive information with NHSmail and other accredited email services.

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

Published by the NIHR Journals Library