

The impact of primary care delays on survival of colorectal cancer patients

Chanpreet Singh Arhi MRCS BSc

Department of Surgery and Cancer
Imperial College London
St Mary's Hospital Campus

Supervisors: Professor Ara Darzi, Mr Paul Ziprin, Professor Paul Aylin

A thesis submitted for the degree of Doctor of Philosophy, December 2018

Declaration of Originality

I declare that this thesis is my own work and, where it is not my own, this is appropriately referenced.

Copyright declaration

The copyright of this thesis rests with the author and is made available under the Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Acknowledgements

Firstly I would like to thank Professor Ara Darzi for providing me the opportunity to embark on this PhD, which has opened the door to other areas of academia. Also I would like to express my gratitude to my co-supervisors, Mr Paul Ziprin and Professor Paul Aylin, who offered the perfect mix between clinical and academic guidance. And thank you to Elaine, George and the BDAU, who provided that initial impetus and the ongoing support to reach this stage.

Four years on, this manuscript is finally written. A lot has happened in that time. Two general elections, one prime minister left, with the other hanging on, just. Brexit looms large but at least the national team provided some welcomed relief.

Yet during these years, the support of my family remained constant, without whom this thesis would never have been completed. Thank you for taking care of our children while I stared endlessly at this screen. In particular to my parents who have provided every opportunity for me to succeed. I hope this thesis goes some way to acknowledging your sacrifices. And to my children who were always willing to offer a welcome distraction.

Finally, I would like to express my deepest appreciation to my wife, Shubha. You have been, and remain still, my strongest advocate. Your unwavering belief has helped through the difficult times when this thesis was impossibly distant.

Contents

DECLARATION OF ORIGINALITY	2
COPYRIGHT DECLARATION	2
ACKNOWLEDGEMENTS.....	3
ABSTRACT.....	8
NOVEL ASPECTS OF THIS RESEARCH	9
LIST OF TABLES.....	10
LIST OF FIGURES.....	13
LIST OF ABBREVIATIONS	15
PUBLICATIONS FROM THIS THESIS	17
PRESENTATIONS FROM THIS THESIS	17
CHAPTER ONE - INTRODUCTION	18
1.1 Hypotheses and aims	18
1.2 Synopsis of studies	19
CHAPTER TWO – BACKGROUND.....	22
2.1 Poor prognosis in the UK	22
2.2 The importance of stage	24
2.2.1 Stage and prognosis	24
2.2.2 Change of stage with patient demographics.....	25
2.2.3 Has the stage at diagnosis changed in the UK over time?	27
2.3 Delays along the pathway for patients diagnosed as a non-emergency	29
2.3.1 Definition of the patient pathway.....	30
2.4 Evidence behind an association between delays and prognosis	40
2.4.1 First symptom to diagnosis or treatment.....	40
2.4.2 Initial presentation to referral.....	41
2.4.3 Referral to diagnosis	42
2.4.4 Initial presentation to diagnosis.....	43
2.4.5 Onset of symptoms to diagnosis	46
2.4.6 Diagnosis to treatment.....	46
2.4.7 Summary of associations between delays and outcome	47

2.5 Emergency diagnosis as a marker of diagnostic delays	48
2.5.1 Why is this important?	49
2.5.2 Definition of an emergency diagnosis	50
2.5.3 Survival following an emergency diagnosis.....	53
2.5.4 Factors associated with an emergency diagnosis	55
2.6 Colorectal cancer delays for young patients	61
2.6.1 Incidence of colorectal cancer in the young	62
2.6.2 Prognosis in young colorectal cancer patients	63
2.6.3 Difference in presentation	66
2.6.4 Summary of young patients and delays in diagnosis	67
2.6.5 Why are young patients defined as under 50 in this thesis?	67
2.7 Weaknesses in previous studies to be addressed in this thesis.....	69
2.8 Datasets used in this thesis – Clinical Practice Research Datalink, the Cancer Registry, Hospital Episodes Statistics and the Office for National Statistics	73
2.8.1 The Clinical Practice Research Datalink.....	73
2.8.2 Hospital Episode Statistics.....	81
2.8.3 Cancer registry	81
2.8.4 Office for National Statistics Mortality data.....	84
CHAPTER THREE - METHODS.....	85
3.1 Extraction of cancer patients from CPRD.....	85
3.2 Definition of symptoms	87
3.3 Patient variables.....	89
3.3.1 Age	89
3.3.2 Co-morbidities.....	90
3.3.3 Socio-economic status	90
3.3.4 GP practice	90
3.4 Tumour variables.....	91
3.4.1 Stage at diagnosis.....	91
3.4.2 Site of cancer	91
3.5 Pathway variables	92
3.5.1 Route to diagnosis.....	92
3.5.2 Definition of referral	93
3.5.3 Type of consultations	93
3.6 Statistical analyses.....	94
3.6.1 Univariate analysis	94
3.6.2 Multivariate analysis	94
CHAPTER FOUR - STUDIES	96
4.1 Study I - Comparison of cancer diagnosis dates between the Clinical Practice Research Datalink, Hospital Episodes Statistics and the Cancer Registry	96
4.1.1 Introduction	97
4.1.2 Method.....	97
4.1.3 Results.....	98
4.1.4 Discussion.....	114

4.2 Study II - Comparison of cancer registry patients from CPRD with national published data from NCIN..	118
4.2.1 Introduction	119
4.2.2 Methods	119
4.3.3 Results	119
4.2.4 Discussion.....	123
4.3 Study III - Complications following colonic resection are associated with a delay in adjuvant chemotherapy treatment which reduces overall survival	125
4.3.1 Introduction	126
4.3.2 Method.....	126
4.3.3 Results	128
4.3.4 Discussion.....	134
4.4 Study IV - Presenting symptoms for non-emergency colorectal cancer patients	139
4.4.1 Introduction	140
4.4.2 Method.....	140
4.4.3 Results	141
4.4.4 Discussion.....	147
4.5 Study V – Are delays in primary care associated with a late stage and worse outcome?.....	150
4.5.1 Introduction	151
4.5.2 Method.....	151
4.5.3 Results	154
4.5.4 Discussion.....	175
4.6 Study VI - Comparison of symptoms and prognosis between two-week-wait and non-urgent referrals.	181
4.6.1 Introduction	182
4.6.2 Method.....	182
4.6.3 Results	183
4.6.4 Discussion.....	193
4.7 Study VII - Missed opportunities in primary care to prevent an emergency diagnosis	197
4.7.1 Introduction	198
4.7.2 Method.....	198
4.7.3 Results	201
4.7.4 Discussion.....	216
4.8 Study VIII – Do young patients experience delays in primary care?	223
4.8.1 Introduction	224
4.8.2 Methods	224
4.8.3 Results	226
4.8.4 Discussion.....	237
CHAPTER FIVE - DISCUSSION	242
5.1 Summary	242
5.2 Implications for clinical practice.....	244
5.3 Comparison to previous studies.....	246
5.4 Limitations.....	247
5.4.1 Missing data	248
5.4.2 Observational rather than randomise controlled study.....	251
5.4.3 Symptoms included in analysis	252
5.4.4 Stage data	252

5.5 Further research	252
5.5 Conclusion	253
REFERENCES	254
APPENDIX	266

Abstract

Survival for colorectal cancer patients in the UK lags behind other Western European countries. This thesis aims to explore the impact of delays in referral and missed opportunities in primary care on survival for colorectal cancer patients. For this purpose patients were identified from the Clinical Practice Research Datalink (CPRD), a national electronic primary care medical record database, with a diagnosis of colorectal cancer in the National Cancer Registry.

The initial studies validate the use of CPRD to fulfil the aims of this thesis, by confirming patients included are representative of the wider national cancer population. In the latter chapters the association of delays in referral with survival are investigated using a retrospective cohort design, while adjusting for patient demographics and co-morbidities. Delays are defined as a referral after three months or after three or more consultations. In addition, missed opportunity for non-emergency diagnosis were identified for emergency patients with a previous consultation in primary care. Survival between the urgent and non-urgent referral pathways are also explored.

This thesis demonstrates delays were associated with a worse survival and later stage for patients presenting with red-flag symptoms (as defined by NICE) but not non-specific symptoms. Presenting with non-specific symptoms was associated with a worse stage and emergency diagnosis independently. There was no difference in survival or stage for patients referred by the non-urgent pathway compared with the two-wait-week pathway.

Identification of cancer patients in primary care is a difficult task. These findings suggest patients presenting with red-flag symptoms, and in particular rectal bleeding are presenting at an earlier stage providing an opportunity for a better prognosis if the referral is prompt. Alternative methods to differentiate cancer from benign patients need to be considered, as relying only on symptoms is insufficient.

Novel aspects of this research

This thesis provides a unique insight into the management of cancer patients in primary care. The studies described here are the first to demonstrate an association between the time to referral and prognosis after a cancer diagnosis. In addition, this thesis describes the delays along the diagnostic pathway dependent on the presenting symptom on a national scale.

List of Tables

Table 1 Interval from the onset of the first symptom to presentation to a primary care physician ³⁵	32
Table 2 Referral delay according to presenting symptom ⁴⁷	36
Table 3 – DOH and NICE referral guidelines for suspected colorectal cancer 2000, 2005 and 2015 ^{49,48,50}	37
Table 4 - Stage according to presenting symptom ⁶³	41
Table 5 - Mortality according to presenting symptom to primary care ⁶²	42
Table 6 - Percentage of patients with cancer related symptoms from four studies	59
Table 7 Prevalence of symptoms in the year leading to colorectal cancer diagnosis	80
Table 8 - Source of cancer diagnosis as defined by the Cancer Registry ¹²⁹	82
Table 8 - Symptoms considered relevant for colorectal cancer for this thesis	87
Table 10 - Comparison of the Routes to Diagnosis and Emergency presentations within 6 months for colorectal cancer cases between 2006 - 2013 inclusive	93
Table 11 – Number of cancer cases recorded in each dataset. N = total number of patients with a cancer diagnosis in at least one of the datasets between 2000 – 2013	99
Table 12 - Comparison of patient demographics and survival for colorectal cancer patients depending on the source	102
Table 13 - Comparison of patient demographics and survival for oesophagogastric cancer patients	103
Table 14 - Comparison of patient demographics and survival for breast cancer patients	104
Table 15 - Comparison of patient demographics and survival for prostate cancer patients	105
Table 16 - Comparison of patient demographics and survival for lung cancer patients depending on	106
Table 17 - Odds ratio for a cancer diagnosis recorded later in CPRD	113
Table 18 Summary statistics comparing patients with first episode adjuvant chemotherapy before and after eight weeks n (%)	130
Table 19 - Complications recorded after colonic resection and before start of adjuvant chemotherapy treatment	131
Table 20 - Odds ratio for first adjuvant chemotherapy treatment after eight weeks or more	132
Table 21 - Hazard ratio of death for patients receiving adjuvant chemotherapy	133
Table 22 - Life table stratified according to first adjuvant chemotherapy episode	134
Table 23 - Demographics and tumour characteristics for 14103 patients with a recording of colorectal cancer in the Cancer Registry between 2000 and 2013	142
Table 24 – Comparison of patient demographics between patients with and without a symptom recorded in the year leading up to cancer diagnosis – univariate and multivariate analysis	144
Table 25 - Percentage of patients presenting with each symptom at the first presentation before a colorectal cancer diagnosis	145
Table 26 The odds ratio of presenting with a particular red-flag symptom, according to age, gender and tumour site and the diagnostic interval for that particular symptom	146
Table 27 – The odds ratio of presenting with a particular non-specific symptom, according to age, gender and tumour site and the diagnostic interval for that particular symptom	147
Table 28 - Odd ratio of a referral according to patient demographics and tumour characteristics. All variables included in the model. Sensitivity analysis for diagnoses from 2006 onwards included	155
Table 29 - Symptom at first consultation in the year leading up to referral before a	156
Table 30 Stage at diagnosis for each presenting symptom, stratified according to site of cancer	158
Table 31 Hazard ratio (HR) of death for colon and rectal cancer patients according to presenting symptom compared with either red-flag symptoms as a group or rectal bleeding only	159
Table 32 - adjusted hazard ratio of death for colon cancer patients according to interval to referral	166
Table 33 - Adjusted hazard ratio of death for rectal cancer patients according to interval to referral	167

Table 34 - Stage at diagnosis for colon cancer stratified according to symptom at presentation	169
Table 35 - Stage at diagnosis for rectal cancer stratified according to presenting symptom	170
Table 36 - Number of consultations before referral for colon cancer patients	171
Table 37 - Number of consultations before referral for rectal cancer patients	171
Table 38 - Median days, and interquartile range, for colon or rectal cancer patients referred after one, two or three or more consultations with relevant symptoms	172
Table 39 - Hazards ratio of death(HR) for a delay in referral (three or more consultations before referral).	175
Table 40 - Patient demographics and tumour characteristics according to the route of diagnosis...	184
Table 41 The referral pathway according to presenting symptom and symptom before referral	185
Table 42 – Referral, hospital and diagnostic interval (median with IQR) and the percentage of patients referred within 2 weeks, 2 weeks to 3 months and over 3 months according to diagnostic route.....	186
Table 43 – Odds ratio (OR) of a two week wait referral based on initial symptom (rectal bleeding as reference) and presence of symptom before referral (absence of symptom as reference).....	187
Table 44 - Demographics and tumour stage according to route of diagnosis for patients with rectal bleeding at first presentation.....	188
Table 45 - Odds ratio of a two week wait referral for patients presenting with rectal bleeding.....	189
Table 46 – Primary care, hospital and total diagnostic interval for TWW and non-urgent patients .	189
Table 47 – Cox regression modelling for survival following referral via two week wait referral, and for all patients dependent on referral delay.....	190
Table 48 – Patient demographics and stage of cancer according to referral route for patients presenting	191
Table 49 – Odds ratio of a Two Week Wait referral for patients presenting with abdominal pain ...	192
Table 50 – Primary care, hospital and total diagnostic interval for patients presenting with abdominal pain	192
Table 51 - Cox regression modelling for survival following referral via two week wait referral, and for all patients dependent on referral delay, for patients presenting with abdominal pain	193
Table 52 – Route to diagnosis for non-emergency colorectal cancer cases between 2006 -13	201
Table 53 - Proportion of patients with an emergency diagnosis according to gender, age, IMD and Charlson score between 2006 – 13 inclusive	202
Table 54 - Presence of a consultation for any reason according to time from diagnosis	203
Table 55 Percentage of patients with relevant consultation in the four years leading up to cancer diagnosis according to non-emergency and emergency routes	204
Table 56 - Presence of symptoms within 28 days of cancer diagnosis.....	206
Table 57 - Presence of symptoms between 28 days and one year before cancer diagnosis-.....	207
Table 58 - Presence of symptoms between one and two years before cancer diagnosis.....	207
Table 59 - Presence of symptoms between two and four years before cancer diagnosis	208
Table 60 The odds ratio for an emergency diagnosis according to patient demographics and presence or absence of symptoms between 28 days and a year before cancer diagnosis	209
Table 61 – The number of colon and rectal cancer patients diagnosed as an emergency (study) and non emergency (control) with a symptom in between 28 days and year before cancer diagnosis...	210
Table 62 - Univariate and multivariate analysis of demographics for colon cancer patients with a symptom recorded between 28 days to one year before cancer diagnosis	211
Table 63 – Univariate and multivariate analysis of demographics for rectal cancer patients with a symptom recorded between 28 days to one year before diagnosis.....	212
Table 64 - Stage at diagnosis for colon and rectal cancer patients for those who had a relevant symptom recorded between 28 days to a year before diagnosis and were diagnosed as a non-emergency (control) and those that had an emergency diagnosis (study).....	212
Table 65 - Symptom before diagnosis for colon cancer patients and OR of the presence of that symptom leading to an emergency diagnosis.....	213

Table 66 - Symptom before diagnosis for rectal cancer patients and OR of the presence of that symptom leading to an emergency diagnosis.....	214
Table 67 - Odds ratio of a late stage (III, IV or missing) for patients with a missed diagnosis according to symptom compared with those diagnosed as a non-emergency.....	215
Table 68 - Patients demographics, stage and site of colorectal cancer	229
Table 69 - Number of consultations in the year leading up to the cancer diagnosis excluding the last 28 days	230
Table 70 - Symptom at presentation after excluding the last 28 days before diagnosis	232
Table 71 - Odds ratio for presentation with a particular symptom	233
Table 72 – Patient demographics and tumour site associated with an Emergency Diagnosis	234
Table 73 - Odds ratio (OR) for emergency presentation dependent on the presenting symptom....	235
Table 74 - The median and interquartile referral, hospital and diagnostic interval in days compared with age.....	236

List of Figures

Figure 1 - Percentage of patients with Stage 1, 2, 3, 4 or unknown colorectal cancer, stratified according to age in England in 2012 ¹⁵	26
Figure 2 - Stage at diagnosis for colorectal cancer between 2000 - 07 ²⁰ in the UK and 2012 - 14 in	27
Figure 3 - Clinical events and time intervals in the patient pathway from first symptom to treatment ²⁹	31
Figure 4 - The positive predictive value for individual or combination of symptoms ⁴²	34
Figure 5 - Percentage of colorectal cancer diagnoses by presentation route from 2006 – 2013. Data source NCIN ⁶⁸	53
Figure 6 - Relative survival by presentation route for colorectal cancer patients diagnosed	54
Figure 7 - European Age-Standardised Incidence Rates per 100,000 Population, ...	62
Figure 8 - European Age-Standardised Mortality Rates per 100,000 Population, by Age, Persons, UK for patient diagnosed with colorectal cancer ¹⁰⁶	64
Figure 9 - Trends in Age-Adjusted US mortality rates by age, Colon and rectum, 1992 - 2014 ¹¹¹	64
Figure 10 - Primary care usage in the two years leading up to cancer diagnosis for benign (lower line) and cancer patients (upper line) ⁴²	88
Figure 11 - Number of patients per year for each of the datasets, stratified according to cancer site	100
Figure 12 - Coverage of cancer diagnoses in source dataset by the other two available dataset for each of the five cancers of interest. NC = No coverage.....	101
Figure 13 - Distribution of differences in diagnosis dates between CPRD and CR.	108
Figure 14 - Distribution of difference in diagnosis dates between CPRD and HES.	109
Figure 15 - Distribution of difference in diagnosis dates between CR and HES.	110
Figure 16 – Number of colorectal cancer cases per year in the linked CPRD cancer registry	120
Figure 17 – Percentage of national cases in England found within the CPRD linked cancer registry	120
Figure 18 – Stage according to year for all colorectal cancer cases in the CPRD linked	121
Figure 19 – Patient demographics between CPRD and the national cancer population in England	122
Figure 20 – Comparison of the route of diagnosis between CPRD and national cancer registry for colorectal cancer diagnoses between 2006 – 13.....	122
Figure 21 – Proportion of cases according to stage, compared between CPRD and the national cancer registry of England 2012-13	123
Figure 22 - Frequency distribution of patients according to first adjuvant chemotherapy episode.....	128
Figure 23 - Kaplan Meier curves for first adjuvant chemotherapy episode.....	134
Figure 24 - Percentage of patients registered with each of the 373 practices	142
Figure 25 - Primary care interval (top left panel), hospital interval (top right panel) and diagnostic interval (bottom panel) in days for colon patients according to presenting symptom	162
Figure 26 Primary care interval (top left panel), hospital interval (top right panel) and diagnostic interval (bottom panel) in days for rectal patients according to presenting symptom	163

Figure 27 - Primary care interval (left), hospital interval (middle) and diagnostic interval (right) for colon (upper row) and	164
Figure 28 - Kaplan Meier demonstrating the difference in survival for colon cancer patients referred after one or two and three or more consultations	174
Figure 29 - Kaplan Meier demonstrating the difference in survival for rectal cancer patients referred after one or two and three or more consultations	174
Figure 30 - Proportion of colorectal cancer diagnoses by age per year from 2000-2013 inclusive	226
Figure 31 - Study patient selection. DCO: Death certificate only IBD: Inflammatory bowel disease	227
Figure 32 - Number of consultations per patient per month stratified according to age group for all consultations and symptom related consultations leading to a non-emergency and emergency diagnosis.....	231
Figure 33 - Comparison of percentage of patients with rectal bleeding or anaemia between my thesis and data from Sheringham et al. ⁹⁴	250

List of Abbreviations

A&E - Accident and Emergency
AC - Adjuvant chemotherapy
AUC - Area Under the Curve ()
CAPER - Cancer Prediction in Exeter
CBH - Change in Bowel Habit
CI - Confidence interval
CPRD - Clinical Practice Research Datalink
CR - Cancer Registry
CRC - Colorectal cancer
CWG - Consensus working group
CWT - Cancer waiting times monitoring
DCO - Death certificate only
DOH - Department of Health
EHR - Excess hazards ratio
EMBASE - Excerpta Medica dataBASE
EMRR - Excess mortality rate ratios
GEE - Generalised Estimating Equation
GI - Gastrointestinal
GP - General Practitioner
GPRD - General Practice Research Database
Hb - Haemoglobin
HES - Hospital Episode Statistics
HR - Hazards ration
HSCIC - Health and Social Care Information Centre
IBM - International Business Machines
ICBP - International Cancer Benchmarking Partnership
ICD - International Classification of Diseases
IMD - Index of multilevel deprivation
IP - Inpatient
IQR - Interquartile Range
IRAC - International Agency for Research on Cancer
ISAC - Independent Scientific Advisory Committee
LR - likelihood ratio
MDT - Multidisciplinary team
MHRA - Medicines and Healthcare products Regulatory Agency
NBCSP - NHS Bowel Cancer Screening Programme
NC - No coverage
NCIN - National Cancer Intelligence Network
NCRAS - National Cancer Registration and Analysis Service

NCWT - National cancer waiting times
NHS - National Health Service
NICE - National Institute for Health and Care Excellence
NIHR - NHS National Institute for Health Research
NS - Not significant
OBH - Other Bowel Habit
OG - Oesophagogastric
ONS - Office For National Statistics
OP - Outpatient
OPCS - Office of Population Censuses and Surveys
OR - Odds ratio
PCR - Primary care record
PHE - Public Health England
PPV - Positive predictive value
QOF - Quality and Outcomes Framework
ROC - Receiver Operating Characteristic
RR - Relative Risk
RS - Relative survival
SEER - Surveillance, Epidemiology and End Results Program
SPSS - Statistical Package for the Social Sciences
TNM staging - Tumour Node Metastasis Classification of Malignant Tumours
TWW - Two week wait
UGI - Upper gastrointestinal
UK - United Kingdom
USA - United States of America
UTS - Up to Standard
Vs - versus
VTE - Venothromboembolic event

Publications from this thesis

Comparison of cancer diagnosis recording between the Clinical Practice Research Datalink, Cancer Registry and Hospital Episodes Statistics

C. Arhi, A. Bottle, E. Burns, J. Clarke, P. Aylin, P. Ziprin, A. Darzi
Cancer Epidemiology December 2018 Vol 57, pp 148-157

Complications after discharge and delays in adjuvant chemotherapy following colonic resection: a cohort study of linked primary and secondary care data

C. Arhi, E. Burns, G. Bouras, P. Aylin, P. Ziprin, A. Darzi
Colorectal Diseases accepted January 2019 – in Press

Presentations from this thesis

Delays in primary care are associated with a late stage and worse prognosis for colorectal cancer patients: A population based study

C. Arhi, E. Burns, A. Bottle, G. Bouras, P. Aylin, P. Ziprin, A. Darzi
NCRI 2018 *British Journal of Cancer* volume 119, pages1–49 (2018)
Bursary Prize winner

Young colorectal cancer patients experience referral delays in primary care leading to emergency diagnoses

C. Arhi, P. Ziprin, A. Bottle, E. Burns, P. Aylin, A. Darzi
NCRI 2018 *British Journal of Cancer* volume 119, pages1–49 (2018)
Bursary Prize winner

Complications following colonic resection delay adjuvant chemotherapy treatment and reduce overall survival: a nationwide study based on routinely collected primary and secondary data

Poster presentation, ACPGBI 2017, 1st author
Published abstract: Colorectal Diseases 19(S4) p14-64

CHAPTER ONE - INTRODUCTION

The clinical pathway for symptomatic cancer starts with the first symptom experienced by the patient¹. During the process of diagnosis and treatment, stakeholders in primary or secondary care aim to reduce delays between key clinical events whilst minimising distress for the patient. This pathway is complex, as a patient may leave and re-enter at any stage². Although there is a focus on identifying cancer patients promptly in primary care, it remains unclear whether current national policies reduce delays or if they have a positive impact on prognosis³. Studies to date have been unable to include sufficient cases to account for important confounders, such as presenting symptoms and patient demographics. In this thesis, the Clinical Practice Research Datalink (CPRD) is used as the source of data. This is a national database of electronic primary care records, which has been matched on an individual basis to Hospital Episodes Statistics (HES), Cancer Registry (CR) and cause of death from the Office for National Statistics (ONS)⁴. The areas of interest for this thesis are

- delays in referral for colorectal cancer patients diagnosed as a non-emergency and the impact on survival
- missed opportunities for non-emergency diagnosis for patients who are eventually diagnosed as an emergency
- delays in diagnosis for young patients compared with older patients

1.1 Hypotheses and aims

The overall hypothesis for this thesis is:

Delays in primary care are associated with a worse outcome for colorectal cancer patients

The aim of this thesis is to answer the following research questions:

- Is CPRD a reliable source to determine delays in cancer diagnosis?
- What is the extent of delays experienced by cancer patients in primary care?
- Are delays in primary care associated with a worse prognosis and/or late stage?
- Is there a difference in outcome between patients referred via the two-week-wait or non-urgent routes?

- Do emergency patients present to their GP before being diagnosed, and if so, why were they not identified?
- Do young patients experience delays in diagnosis compared with the older age groups?

1.2 Synopsis of studies

The following studies have been carried out to fulfil the above aims. The hypothesis and the aims of each study are expanded upon in the relevant chapters.

Study I - Comparison of cancer diagnosis dates between the Clinical Practice Research Datalink, Hospital Episodes Statistics and the Cancer Registry

Hypothesis: Relying solely on CPRD to identify cancer cases leads to a selection bias by excluding certain patient demographics

Cancer diagnoses in CPRD are recorded later than the cancer registry and HES

Aims: Compare the completeness and the date of diagnosis between three available national datasets to determine the preferred source of cancer patients

Study II - Comparison of cancer registry patients from CPRD with national published data from the National Cancer Intelligence Network

Hypothesis: Patients extracted from the CPRD linked cancer registry are representative of the national population

Aims: Compare the demographics of patients and stage identified from the CPRD linked cancer registry with national cancer registry data

Study III - Complications following colonic resection are associated with a delay in adjuvant chemotherapy treatment which reduces overall survival

Hypothesis: Complication following a resection for colon cancer is associated with a delay in adjuvant chemotherapy and worse prognosis

Results from CPRD concerning delays are comparable to other national datasets

Aims: Compare the results regarding delays to other sources, thereby adding further validity to the use of CPRD to investigate delays

Study IV - Presenting symptoms for non-emergency colorectal cancer patients

Hypothesis: The symptom at presentation for non-emergency patients varies according to patient demographics

Aims: To improve our understanding of the presenting symptom according to gender, age and co-morbidities and how this correlates with national referral guidelines

Study V - Are delays in primary care associated with a late stage and worse outcome?

Hypothesis: Delays in primary care are associated with a later stage and worse outcome for colorectal cancer patients

Aims: Determine the primary care (pre-referral), hospital (referral to diagnosis) and whole diagnostic interval according to presenting symptom
Determine the patient demographics most likely to lead to a delay in referral
Determine whether delays in primary care are associated with a later stage and worse prognosis

Study VI - Comparison of symptoms and outcome between urgent and non-urgent referrals

Hypothesis: Patients diagnosed via the urgent (two-week-wait pathway) have a better prognosis than those referred by the non-urgent pathway

Aims: Determine the patient characteristics and symptoms most likely to lead to an urgent referral from primary care
Investigate whether there is a difference in stage or prognosis between the two pathways after accounting for presenting symptoms

Study VII - Are there missed opportunities in primary care to prevent an emergency diagnosis?

Hypothesis: Patients who are diagnosed as an emergency experience missed opportunities in primary care for a non-emergency diagnosis

Aims: Describe the patient characteristics and symptoms most likely to lead to a missed opportunity for a non-emergency diagnosis

Determine whether there is a difference in stage at diagnosis between emergency and non-emergency patients despite presenting with the same symptoms

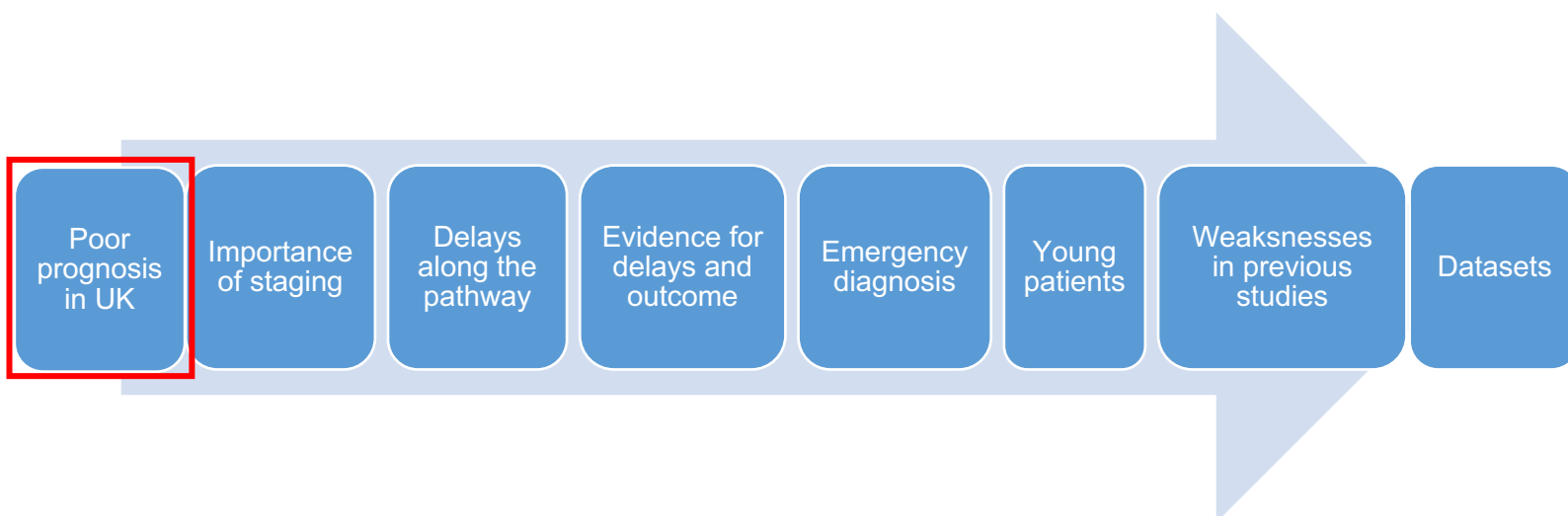
Study VIII - Do young patients experience delays in primary care?

Hypothesis: Young patients with colorectal cancer (defined as those aged 50 or under) experience delays in diagnosis in primary care

Aims: Determine the differences, in the presenting symptoms of young versus older patients

Investigate the risk of an emergency diagnosis for young patients even if they were to present with the same symptoms to their GP as older patients

CHAPTER TWO – BACKGROUND



2.1 Poor prognosis in the UK

In 2000 the Department of Health in the UK released guidelines to aid General Practitioners (GP) in identifying patients with suspected cancer⁵. These criteria were in response to the relatively poor outcome in the UK following cancer diagnosis compared to western European countries. The first Eurocare study⁶ used cancer registry data from the period 1979 – 1984, to demonstrate the relative risk of excess death for colorectal cancer in the UK (represented by Thames, West Midlands and Yorkshire) was 1.58 (95% CI 1.39 - 1.79) in the first 6 months after diagnosis, and 1.27 (95% CI 1.03 – 1.57) subsequently up to five years, compared with Finland. If survival in the UK was similar to the European mean it has been postulated that over 800 colon cancer, 200 rectal cancer and 600 stomach cancer related deaths within 5 years of diagnosis could have been avoided between 1995 - 1999⁷. The latest ICBP study published in 2019 has demonstrated the survival following colon or rectal cancer diagnosis in the UK and Denmark has improved the most from 1995 to 2014, but still remains lower than the other high income countries included in the study⁸. The following reasons have been suggested to explain the relatively poor outcome in the UK.

1. *Quality of UK registry data*

Since the inception of the initial Eurocare study, the UK cancer registries have adopted guidelines similar to those set out by the International Agency for Research on Cancer^{9,10}. Subsequent cancer registry analysis, with close to 100% UK coverage in the Eurocare-5 study, described a 5-year-age-standardised relative survival of 51.3% (95% CI 51.0 – 51.7) for

colon cancer, 53.7% (95% CI 53.3 – 54.2) for rectal cancer and 17.0% (95% CI 16.6 – 17.4) for stomach cancer¹¹. Apart from Denmark, these results compared poorly with all other Northern and Central European countries included in the study, suggesting survival following diagnosis in the UK could be improved. These findings have been confirmed independently by the International Cancer Benchmarking Partnership¹², a global group aimed to identify reasons for deficits in cancer survival, and the CONCORD working group¹³. Therefore the quality of registry data can no longer be advocated as a reason for differences in outcome.

2. Differences in methods of resection and post-surgical management

Although there are differences in treatment between countries, the evidence this is a significant cause is less convincing and is outside the scope of this thesis. Nevertheless it is understandable that countries where patients are diagnosed at an earlier stage would have a higher rate of resections with curative intent.

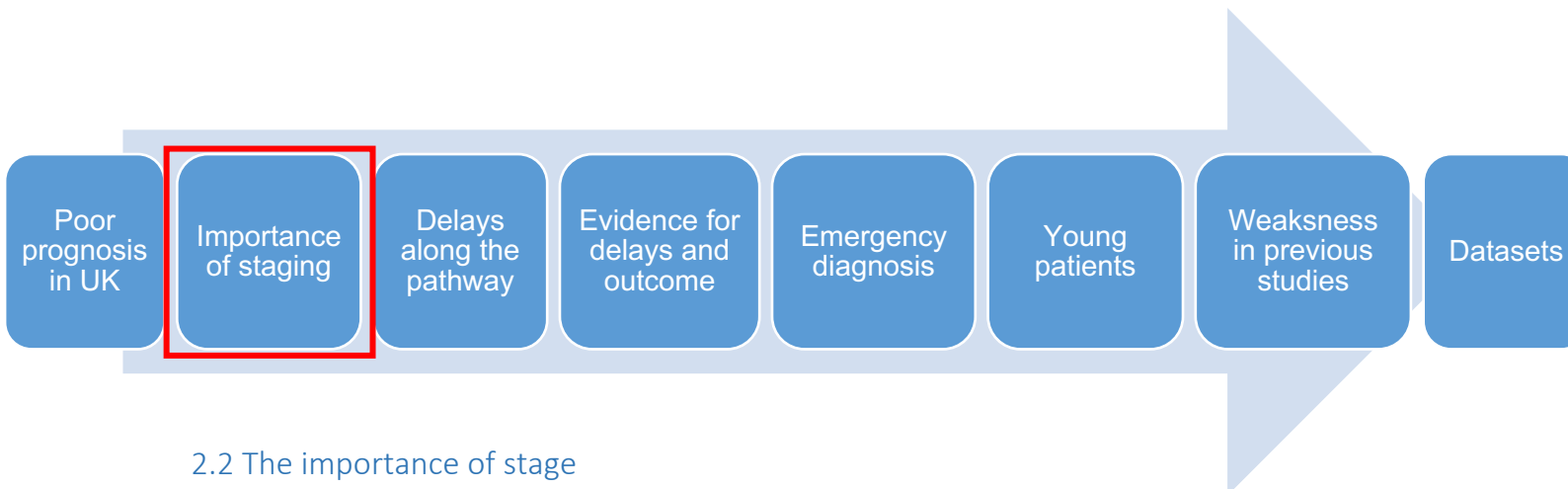
3. Differences in healthcare systems

In the UK, as in Denmark, the general practitioner acts as a gate-keeper for further investigations and secondary care, which does not require payment from the patient¹⁴. As a result waiting times to be seen by a specialist are long. This has led to national referral criteria based on symptoms and blood results in both countries that are designed to reduce delays for specialist input and/or diagnostics. In the countries with better survival, such as Sweden, patients can bypass primary care physicians and present directly to hospitals¹⁴. These differences suggest delays along the primary care pathway should be investigated as a cause of the worse outcome in the UK.

4. Delays along patient pathway

This potential factor forms the emphasis of this thesis. Delays in diagnosis have been suggested as cause of a later stage at diagnosis for non-emergency patients and a higher proportion of emergency diagnosis. The latter is independently associated with a late stage diagnosis. Before discussing the factors associated with delays along the diagnostic

pathway, the importance of an earlier stage and how this differs according to patient demographics and country is described below.



2.2 The importance of stage

2.2.1 Stage and prognosis

Stage at diagnosis is a key determinant for patient outcome^{15,16}. Analysis of colorectal cancer cases diagnosed from 1996 to 2006 demonstrated a fall in 5 year relative survival from 93.2% (95% CI 92.5 – 93.9) for stage A to 6.6% (95% CI 6.1 – 7.0) for stage D¹⁵. The difference in survival is less pronounced in the first year after diagnosis: Relative survival for Stage A was 95.7% (95% CI 95.3 – 96.2) vs stage D 34.2% (95% CI 33.7 – 34.7) for colon cancer¹⁶. (Relative survival defined here as the ratio of the observed probability of survival and the probability that would have been expected had the patients experienced the normal (background) mortality of the population in which they live, given the same distribution of factors such as age, sex and calendar period). Therefore it would be reasonable to attempt to diagnose patients at an earlier stage of their disease. This forms the basis of screening programmes¹⁷.

Stage data from EURO CARE-1 study suggested that patients were diagnosed at a later stage in the UK compared to better performing European countries¹⁸. This same data when compared to the United States, found similar results¹⁹. However the UK was represented by two UK registries with relatively poor stage ascertainment²⁰. Maringe et al.²¹ used colorectal cancer stage data from 2000 - 2007 and found that the proportion of patients with advanced stage (Dukes C or D) was low in the UK (38% of all cases including missing),

but had the highest proportion of missing stage (27.8%), leading to survival calculated based on imputed data. The ICBP^{12,22} demonstrated that the difference in one year survival between the UK and Sweden (74.7% vs 83.8% respectively), one of Europe's better performing countries, was larger for colorectal cancer than the difference in the conditional 5 year survival (71.8% vs 74.8%). The conditional 5 year survival is defined as the proportion of patients that survive to 5 years, after the exclusion of patients who died within a year of diagnosis. This measure potentially describes the effect of treatment, while the one year survival reflects the stage at diagnosis. These findings were confirmed in the Eurocare-4 study, that had complete registration data for England and the other three countries of the UK^{23,24}.

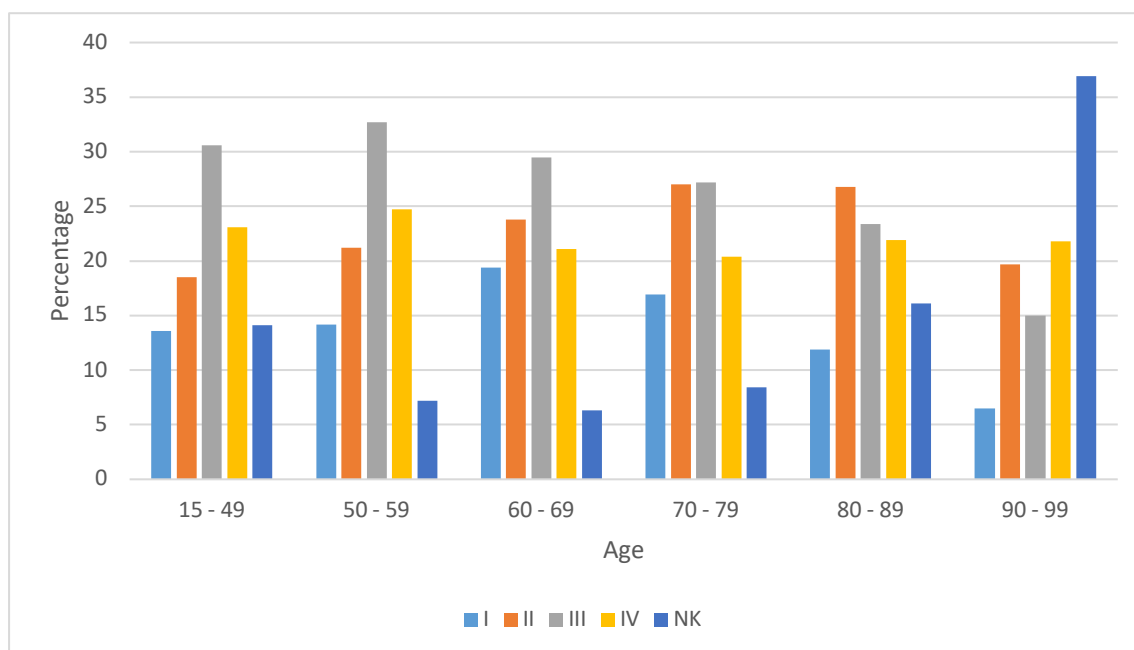
2.2.2 Change of stage with patient demographics

The stage at diagnosis varies with patient demographics. McPhail et al¹⁶. analysed all colorectal cancer, including emergency (in addition to breast, prostate, lung and ovarian cancer) in 2012 in England. The authors compared stage at diagnosis according to age, gender and deprivation index, as well as relative survival and excess mortality. However co-morbidities were not included. The findings are summarised below:

- In 2012 there was 89% complete stage data for 34011 colorectal cancer patients, consisting of 15.5% Stage 1, 24.7% Stage 2, 27.2% Stage 3 and 21.6% stage 4 (figure 1).
- There is a small variation in stage 1 at diagnosis between male and female patients (16% vs 14% respectively), but there is no significant difference between the genders
- Patients aged 60 – 69 have the highest proportion of stage 1 cancer (19.4%). This compares to 13.6% of those aged 15-49 and 6.5% of those aged 90 – 99.
- 30.6% of 18-49 year olds and 32.7% of 50 – 59 have stage 3 cancer. The remaining age groups have a lower proportion of stage 3
- 23.1% of 18-49 year olds and 24.7% of 50 – 59 have stage 4 cancer. The remaining age groups have a lower proportion of stage 4
- Unknown stage is highest for the oldest and youngest groups
- Stage at diagnosis is higher for the most deprived income groups, although the differences are marginal

- Adjusted relative survival (comparison to the expected survival in the general population matched for age, gender, sociodemographic status and geographical region) decreases with stage: 98.0% stage 1, 94.6% stage 2, 89.1% stage 3, 43.8% stage 4 and 60.2% for unknown stage
- Excess mortality rate ratios (EMRR), modelled using gender, age, income deprivation quintile, geographical location and stage at diagnosis as independent variables, was also calculated. Female patients were at an increased risk of death (EMRR 1.14 95% CI 1.08-1.19 $p < 0.01$) compared with male patients. Increasing age also increased the EMRR from 1.3 (95% CI 1.12-1.52) for 50-59 year olds to 1.73 (95% CI 1.50-1.98) for those aged 60-69 compared with 15-49 year olds. Increasing level of deprivation independently increased the EMRR (1.39 95% CI 1.28-1.50) for the most deprived quintile compared with the least deprived. Worsening stage, after taking into account the above factors, increased the EMRR.

Figure 1 - Percentage of patients with Stage 1, 2, 3, 4 or unknown colorectal cancer, stratified according to age in England in 2012¹⁶

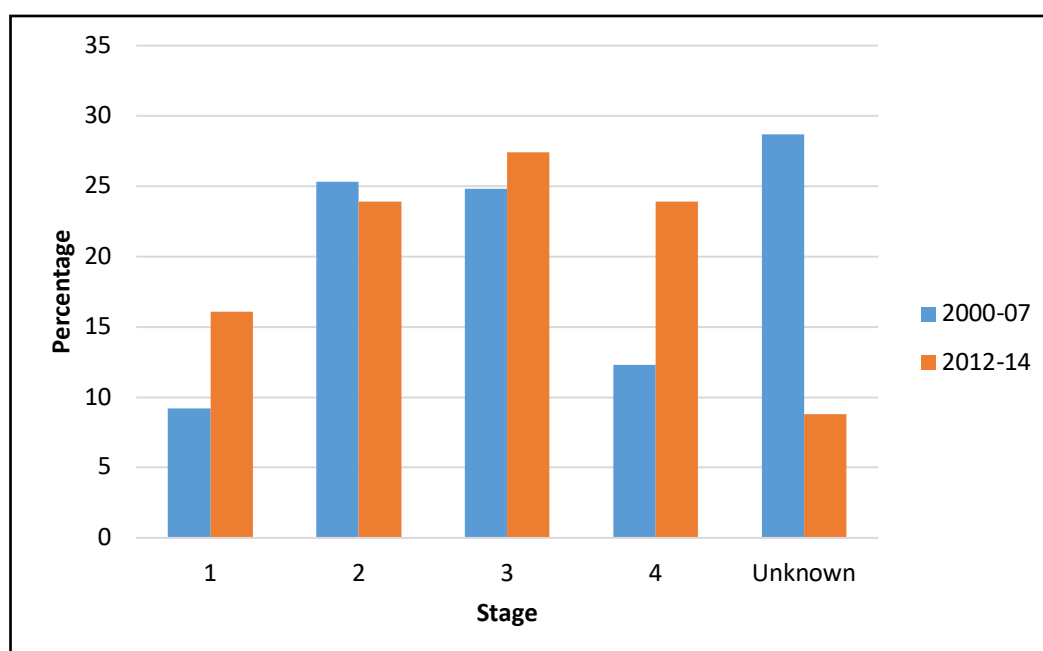


This study suggests that age exerts a strong influence on stage at diagnosis, although gender and deprivation index are not significant factors. The youngest age demonstrated the best survival despite being diagnosed at a later stage. As all these factors determine prognosis, they should be included as independent variables in models of survival.

2.2.3 Has the stage at diagnosis changed in the UK over time?

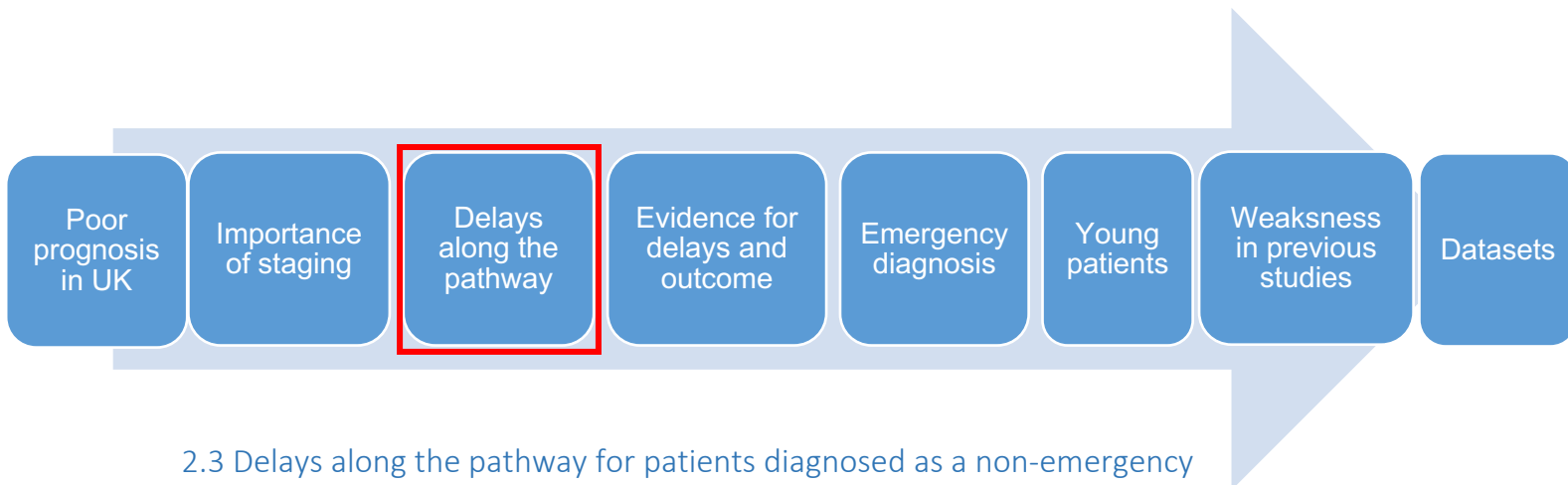
Over the last 15 years, there has been an improvement in the stage at diagnosis for colorectal cancer (figure 2) in the UK, and its constituent countries. Between 2000 – 07²¹ and 2012 – 14²⁵, there was a fall in cases with unknown stage from 28.7% to 8.8%, reflecting the improvement in accuracy by individual cancer registries and better co-ordination of data collection at a national level. There was an increase in stage 1 disease from 9.2% to 16.1%, but also an increase in stage IV from 12.3% to 23.9%. The former may be due to the introduction of NICE guidelines and screening. The latter indicates previous patients who would have been classified as ‘missing stage’, perhaps as they died soon after diagnosis, may now be correctly classified with stage IV disease. Studies have shown the proportion of stage IV disease increases after imputation²¹. It is important to note the 2000 - 07 data is from the analysis of cancer registry data from the UK²¹, while the later years data has been provided by Public Health England²⁶ so includes only data from England.

Figure 2 - Stage at diagnosis for colorectal cancer between 2000 - 07²¹ in the UK and 2012 - 14 in England for all patients²⁶



Data from Public Health England²⁶ has also shown an improvement in the relative 1 year survival following colorectal cancer diagnosis from 72.4% (95% CI 72.2 – 72.7) in 2001 –

2005 to 78.1% (95% CI 77.9 – 78.3) in 2009-13 and 5 year survival from 50.8% (95% CI 50.5 – 51.2) to 58.9% (95% CI 58.6 – 59.2) respectively. These figures have been adjusted for gender and age. However there still remains scope to improve the stage at diagnosis, especially for the younger age groups. Such an improvement may be achieved by reducing delays in diagnosis²⁷. In the following chapter I define the intervals along the patient pathway, and the patient and tumour factors associated with delays, together with the current evidence regarding associations between delays and prognosis.



For non-emergency colorectal cancer patients, I aim to investigate two hypotheses:

Delays in referral from primary care are associated with a late stage of colorectal cancer and a worse prognosis

The aims of this study are to answer the following questions:

- Which presenting symptoms are associated with delays in referral from primary care?
- How does the hospital interval differ according to the presenting symptom?
- Does the presenting symptom alter the stage at diagnosis?
- Are delays in primary care associated with a later stage of cancer?
- Are delays in primary care associated with a worse prognosis?

Patients referred via the two-week-wait pathway demonstrate a significantly better prognosis than patients referred via the non-urgent pathway

I aim to answer the following research questions:

- What is the difference in patient demographics and stage between the two pathways?
- Which symptoms are most likely to lead to a non-urgent referral?
- Does the hospital interval differ according to the referral pathway?

- After adjusting for patient demographics, the presenting symptom and the hospital interval, is there a difference in prognosis between the two pathways?

Patients diagnosed as an emergency are discussed in the next chapter

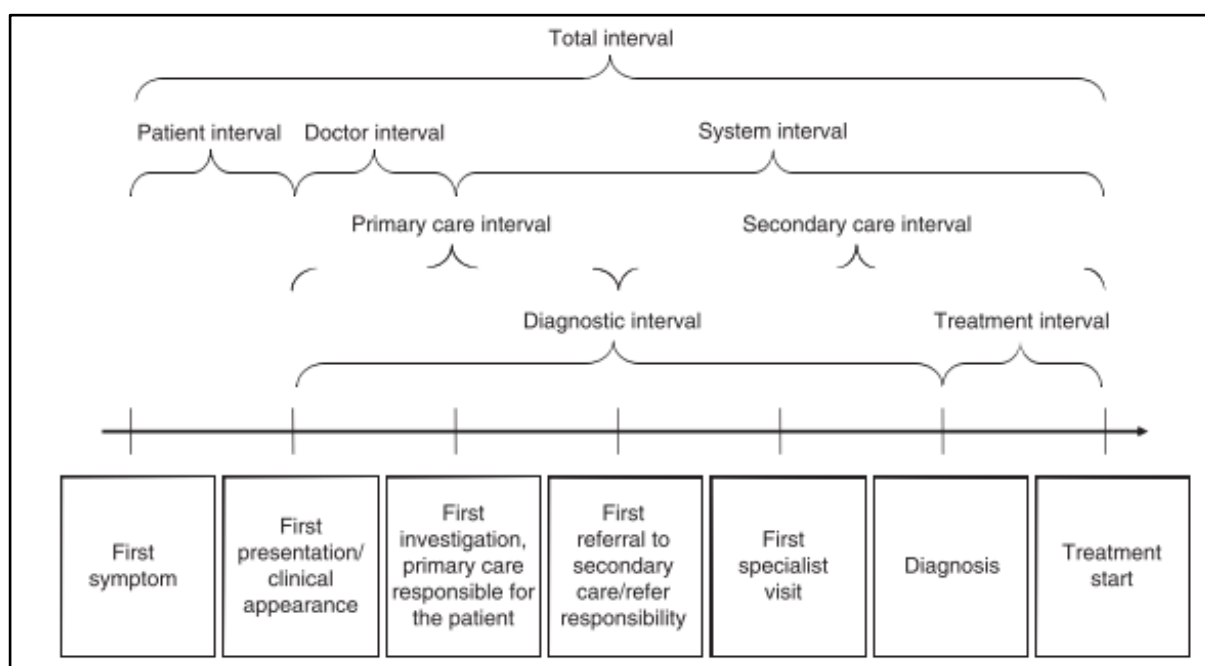
2.3.1 Definition of the patient pathway

Despite the national screening programmes, over 90% of cancer patients are diagnosed following the onset of symptoms^{28,29}. If this prompts health seeking behaviour, the patient may either present to their primary care physician (General Practitioner in the UK) or the emergency department. By choosing the GP as the first point of contact, the current pathway in the UK follows through similar steps as described by Olesen et al.³⁰ (figure 3). There are a number of potential bottlenecks along this process, created by patients themselves, the doctors they consult or the primary and secondary care system. For my thesis I concentrate on the interval between first presentation and diagnosis, with some investigation into the interval to treatment.

The representation of the patient pathway by Olesen et al.³⁰ was adopted by the Consensus Working Group (CWG)¹, a group consisting of experts in the field of early cancer diagnosis, commissioned by Cancer Research UK and the Department of Health in England. The CWG published the Aarhus statement¹, which provides definitions of key clinical events along the patient pathway and methodological approaches when carrying out research involving the diagnosis of cancer. I have followed these definitions in this thesis.

Factors associated with a delay in patient or diagnostic interval have been described in three comprehensive systematic reviews^{31–33}. I have used these systematic reviews together with more recent studies to describe the factors associated with a delay. A formal systematic review was not necessary.

Figure 3 - Clinical events and time intervals in the patient pathway from first symptom to treatment³⁰



2.3.1.1 Patient interval - from first symptom to first contact with the GP

The start of this interval is defined as the ‘time point when first bodily changes and/or symptoms are noticed’¹. Although I have not investigated associations between delays during this interval and outcome, it is important to understand the potential contribution this interval may make to the overall diagnostic interval.

The reasons a patient initiates health seeking behaviour is complex. A patient may attribute the new symptom to a benign cause³⁴ and so may self-manage, expecting the symptom to resolve³³. This is more likely for vague, non-specific symptoms³² or for symptoms which have a well-known non-malignant condition. For example rectal bleeding may lead to a delay in presentation for colorectal cancer³² as the patient assumes it is due to haemorrhoids. On the contrary upper GI bleeding has been shown to reduce the patient interval for upper gastrointestinal cancer³¹ as this is considered an unusual and concerning

symptom. Knowledge of the symptoms associated with cancer can reduce this interval³⁵. Patients who perceive their symptoms as 'more severe' are likely to present with less delay³¹. A similar finding was found for those with multiple symptoms. Recurrence of symptoms or a previous history of cancer also reduces delay. Walter et al.³⁶ described the patient interval according to presenting symptom, for all cases referred to secondary care with suspected colorectal cancer (table 1). Although this study showed rectal bleeding had the shortest patient interval, there was not a significant difference between symptoms.

Table 1 Interval from the onset of the first symptom to presentation to a primary care physician³⁶

	Median (IQR) days
Any symptom	35 (7–92)
Change in bowel habit	42 (14–92)
'Bleeding from back passage'	25 (3–69)
Indigestion/heartburn/persistent tummy ache	31 (5–86)
Fatigue or tiredness	31 (7 - 89)
Feeling different 'in yourself' from usual	30 (4–76)
Decrease in appetite	20 (3–62)
Back pain	29 (2–76)
Unexplained weight loss	31 (5–92)

Patients of a lower socioeconomic class were less likely to recognise a change in bowel habit. The patients ethnic background plays a role in the way the patient manages their symptoms³⁴, with non-white patients less likely to delay for stomach cancer³¹. Although older age has been found to be a risk factor for delayed presentation for breast cancer³⁷, there is no evidence for association between age and delay for other cancers. There is no strong evidence for an association between marital status or gender and presentation delay for colorectal cancers^{32,33}. Patients with more co-morbidities were associated with a reduced delay for colorectal cancer³². The fear of investigations, or of the diagnosis itself, can delay presentation^{34,35,38}. Rural setting and lower level of education increased delays.

Within this interval, a delay may be caused by waiting for an appointment with the GP, potentially encouraging the patient to seek medical attention directly with secondary care via the emergency department.

2.3.1.2 Primary care interval – from first presentation to referral

This interval forms the main emphasis of my thesis. I define the start of this interval as the first consultation with the GP with a symptom that could be attributed to the cancer in question. The end of the interval is the date the GP refers to secondary care for a suspected cancer¹. Misdiagnosis that the presenting symptoms are due to an underlying benign condition is the main reason why there may be a delay in referral^{31,32}. The factors that may influence this decision can be grouped as presenting symptom, practitioner behaviour and patient factors.

Two recent large national audits of primary care practice have been produced, one based on a GP questionnaire³⁹, and the other on a patient survey⁴⁰. Neither of these data were stratified according to presenting symptoms. Based on the GP data, colorectal cancer patients experienced an interval of median 7 days (IQR 0 – 32) from presentation to referral and 21.3% of patients were referred after three or more consultations. This latter figure increased to 42.8% from the patient survey for colon cancer and 22.5% for rectal cancer. According to the patient survey, 42.5% of colon cancer and 55.1% of rectal cancer patients were referred after the first consultation, perhaps reflecting increased likelihood rectal cancer presents with rectal bleeding and/or palpable mass on rectal examination.

2.3.1.2.1 Presenting symptom

When a patient first presents to their GP, it can be a challenge to distinguish a patient with cancer from the vast majority that have an underlying benign condition as the range of symptoms is similar². A full-time GP will only see one or two cancer patients diagnosed per year with one of the four more common cancers – breast, colon, lung and prostate⁴¹. In my thesis I will be analysing delays based on the individual symptoms associated with colorectal cancer. For this purpose, it is necessary to understand which symptoms should be included in my analysis. There is a paucity of studies that have described the percentage of all patients, whether due to cancer or a benign condition, who present to their primary care physician with particular symptoms.

The most common measurement used to describe the difference in symptoms between cancer and benign patients are the positive predictive value (PPV) and the likelihood ratio

(LR) of a single or combination of symptoms. The PPV is defined as the proportion of patients with a positive test who actually have the disease. This test is dependent on the prevalence of disease in the population tested, in contrary to the LR which is independent⁴².

The PPV increases further along the pathway as patients without cancer are filtered out based on clinical findings or investigations: the community, in primary care, in outpatients clinic or at endoscopy. Hamilton et al.⁴¹ described the PPV for symptoms recorded in the medical notes for patients who presented to their GP by matching cancer cases with five controls based on gender, general practice and age (figure 4). A criticism of this study is that cancer cases were match-controlled to a limit of 5:1 patients to calculate the PPV. This metric should be measured after consideration of all patients who attend primary care to provide a true PPV. This weakness in the study design precluded it from consideration for NICE guidance.

Figure 4 - The positive predictive value for individual or combination of symptoms⁴³

Constipation	Diarrhoea	Rectal bleeding	Loss of weight	Abdominal pain	Abdominal tenderness	Abnormal rectal exam	Haemoglobin 10–13 g dl ⁻¹	Haemoglobin <10 g dl ⁻¹	
0.42 0.3, 0.5	0.94 0.7, 1.1	2.4 1.9, 3.2	1.2 0.9, 1.6	1.1 0.9, 1.3	1.1 0.8, 1.5	1.5 1.0, 2.2	0.97 0.8, 1.3	2.3 1.6, 3.1	PPV as a single symptom
0.81 0.5, 1.3	1.1 0.6, 1.8	2.4 1.4, 4.4	3.0 1.7, 5.4	1.5 1.0, 2.2	1.7 0.9, 3.4	2.6 0.6, 2.7	1.2 0.6, 2.7	2.6	Constipation
	1.5 1.0, 2.2	3.4 2.1, 6.0	3.1 1.8, 5.5	1.9 1.4, 2.7	2.4 1.3, 4.8	11 1.2, 4.3	2.2 1.2, 4.3	2.9	Diarrhoea
		6.8 1.9, 5.3	4.7 1.9, 5.3	3.1 1.4, 5.2	4.5 1.3, 5.2	8.5 1.3, 5.2	3.6 1.1, 4.5	3.2	Rectal bleeding
			1.4 0.8, 2.6	3.4 2.1, 6.0	6.4 0.3, 5.2	7.4 0.3, 5.2	1.3 0.7, 2.6	4.7	Loss of weight
				3.0 1.8, 5.2	1.4 0.3, 5.2	3.3 0.3, 5.2	2.2 1.1, 4.5	6.9	Abdominal pain
					1.7 0.8, 3.7	5.8 0.8, 3.7	2.7 0.8, 3.7	>10	Abdominal tenderness

The top line represents the PPV of an individual symptom. Combination of the same symptom is the PPV if a patient presents with the same symptom more than once. The remaining cells describe the PPV for combination of symptoms. The yellow shading is when the PPV is above 1%, the amber when its over 2.5% which approximates to a colorectal cancer risk of over 10 times normal. The red is when the PPV is over 5% and approximate a risk of over 20 times normal

In a study by Walter et al,³⁶ patients over the age of 40 were recruited after referral to secondary care, thereby excluding patients who were managed solely in primary care. This study found rectal bleeding was the only symptom that colorectal cancer patients were significantly more likely to experience than patients with an underlying benign condition (n

= 52 (34.2%) vs n = 552 (23.9%) respectively p = 0.004). The proportion of patients with a change in bowel habit, abdominal pain, weight loss or fatigue were similar between the two groups.

Hippisley-Cox et al.⁴⁴, derived the following PPV from a study cohort of 2 351 052 patients: Rectal bleeding 2.9%, Abdominal pain 0.7%, weight loss 0.8%, change in bowel habit 2.8%, anaemia 1.5%. After adjusting for age, alcohol status and family history of gastrointestinal cancer, they found rectal bleeding (HR 27.0 95% CI 23.5 to 31.1), abdominal pain (HR 6.78 95% CI 5.76 to 7.97), loss of appetite (HR 2.15 95% CI 1.53 to 3.03), weight loss (HR 4.07 95% CI 3.42 to 4.85) and anaemia (HR 3.33 95% CI 2.87 to 3.87) were all associated with an increased risk of colorectal cancer for males and females (the HR for females are of similar magnitude as those described for male patients).

In contrast a study by Panzuto et al.⁴⁵ based in the primary care setting in Italy demonstrated a higher PPV for these symptoms: Abdominal pain 3.3% (95% CI 0.69 – 15.6), anaemia 9.7% (95% CI 3.52 to 26.8), weight loss 35.7% (95% CI 9.3 to 18.6), change in bowel habit 14.0% (95% CI 6.26 to 25.8), diarrhoea 11.8% (95% CI 5.8 to 20.6), constipation 15.7% (95% CI 10.0 to 23.0) and bloating 13.2% (95% CI 8.44 to 19.3). The difference in findings is probably because the primary care physicians involved in the Panzuto et al.⁴⁵ study underwent a training course about the possible symptoms associated with colorectal cancer before the collection of data.

In a systematic review by Ford et al.⁴⁶ the LR for individual symptoms was based on studies set in secondary care so was not a true representation of the challenges faced by GPs. In a systematic review, Astin et al.⁴⁷ carried out a pooled analysis of only studies based in primary care, but did include the Hamilton et al.⁴¹ data mentioned above. This review noted there was a high degree of heterogeneity between the studies. Astin et al.⁴⁷ found the PPV for rectal bleeding alone was 8.1% (95% CI 6.0 – 10.8), with abdominal pain 7.58% (95% CI 3.0 – 19.2), with change in bowel habit 11.8% (95% CI 6.78 – 20.4), with anaemia 21.4% (95% CI 4.70 – 50.8), with weight loss 13.4% (95% CI 8.15 – 21.9), with diarrhoea 7.4% (95% CI 0.91 – 24.3) and with constipation 2.6% (95% CI 0.07 to 3.5).

Barrett et al.⁴⁸, from a study in 2002 based on the original CAPER study, demonstrated patients presenting with rectal bleeding, abdominal mass or diarrhoea had the shorter referral interval than those with anaemia, constipation or abdominal pain (table 2). There was no investigation as to whether delays in primary care were associated with a worse prognosis.

Table 2 Referral delay according to presenting symptom⁴⁸

	n	Median (IQR)
Abdominal mass	19	0 (0,1)
Rectal Bleeding	47	2 (0, 7)
Diarrhoea	33	4 (0, 44)
Abdominal pain	45	5 (0, 69)
Constipation	27	7 (1, 182)
Anaemia	38	11 (0, 54)
Positive faecal occult blood test	11	14 (8, 23)

Despite certain weaknesses in these studies, it is possible to conclude a GP faces a difficult task to differentiate the very few patients whose symptoms are due to cancer. To assist the GP, the Department of Health in 2000⁴¹, followed by NICE in 2005⁴⁹ and again in 2015⁵⁰, issued national guidelines regarding the symptoms that warrant an urgent referral to secondary care (table 3). Each iteration of the guidelines progressively reduced the threshold for referral. The initial guidance in 2000 was based on expert consensus. The 2005 guidelines were similar in scope to the 2000 guidance as there was insufficient evidence to significantly alter the criteria. However the 2015 version took into account a plethora of new studies that allowed a more detailed guidance based on age and symptom combinations. The latest guidance was the first to use a threshold for the positive predictive value for symptoms, which was set at 3%⁵⁰.

Table 3 – DOH and NICE referral guidelines for suspected colorectal cancer 2000, 2005 and 2015⁵⁰⁴⁹⁵¹

2000 - DOH	2005 - NICE	2015 - NICE
Rectal bleeding with a change in bowel habit to looser stools and/or increased frequency of defecation persistent for 6 weeks, irrespective of age	Aged 40 years and older, rectal bleeding with a change of bowel habit towards looser stools and/or increased stool frequency persisting for 6 weeks or more	Aged 50 and over with unexplained rectal bleeding
Aged 60 years and older with rectal bleeding persistently without anal symptoms	Aged 60 years and older, with rectal bleeding persisting for 6 weeks or more without a change in bowel habit and without anal symptoms	Under 50 with rectal bleeding and any of the following unexplained symptoms or findings: ▪ abdominal pain ▪ change in bowel habit ▪ weight loss iron-deficiency anaemia
Aged 60 years and older with a change of bowel habit to looser stools and/or increased frequency of defecation, without rectal bleeding and persistent for six weeks	Aged 60 years and older, with a change in bowel habit to looser stools and/or more frequent stools persisting for 6 weeks or more without rectal bleeding	Aged 60 and over with changes in their bowel habit
A definite palpable right-sided or rectal mass	Right lower abdominal or rectal mass	Rectal or abdominal mass
Iron deficiency anaemia without an obvious cause (Hb < 11 g/dl in men or < 10 g/dl in postmenopausal women)	Iron deficiency anaemia without an obvious cause (Hb < 11 g/dl in men or < 10 g/dl in postmenopausal women)	Aged 60 and over with iron-deficiency anaemia
		Aged 40 and over with unexplained weight loss and abdominal pain
		Tests show occult blood in their faeces

The review by Mitchell et al.³² stated that it was difficult to make a conclusion about how the symptom at presentation influenced the speed of referral. Some studies suggested colorectal cancer patients with rectal bleeding had experienced a delay as their symptom was initially attributed to haemorrhoids, although other studies demonstrated rectal bleeding reduced the referral and diagnostic process. Similarly experiencing pain resulted in more rapid referral for some patients, whereas for others this had no impact. The severity and chronicity of the pain could make a difference, as patients with non-severe abdominal

pain were more likely to be investigated in primary care. Rectal cancer patients were less likely to experience a primary care delay compared to colon cancer patients.

Many of the studies included in the review by Mitchell et al.³² were carried out before the introduction of the national referral guidelines. Since their implementation, patients who fulfil the criteria are referred and diagnosed quicker than previously, but patients who fall outside this criteria due to non-specific symptoms at presentation, which can be over 50%, have a longer waiting time to diagnosis⁵².

2.3.1.2.2 Practitioner behavior

In the UK, like in Denmark, the GP plays an important 'gate-keeper' role where access to secondary care requires a referral from the GP⁵³. Both these countries are associated with poor cancer outcome, and so the structure of the healthcare system and the GP themselves¹⁴, have been identified as possible causes for delays in diagnosis. As such, 31.7% of patients with colorectal, 36.0% of stomach and 24.9% of oesophageal cancer patients, experienced three or more consultations before referral in an audit of English practices³⁹. Although more recent data upto 2014 has shown the survival gap between Denmark and the UK compared with high earning countries in Europe and around the world has narrowed, it still remains lower (5 year survival has improved by over 13% in both these countries between 1995 and 2014 for colon and rectal cancer)⁸.

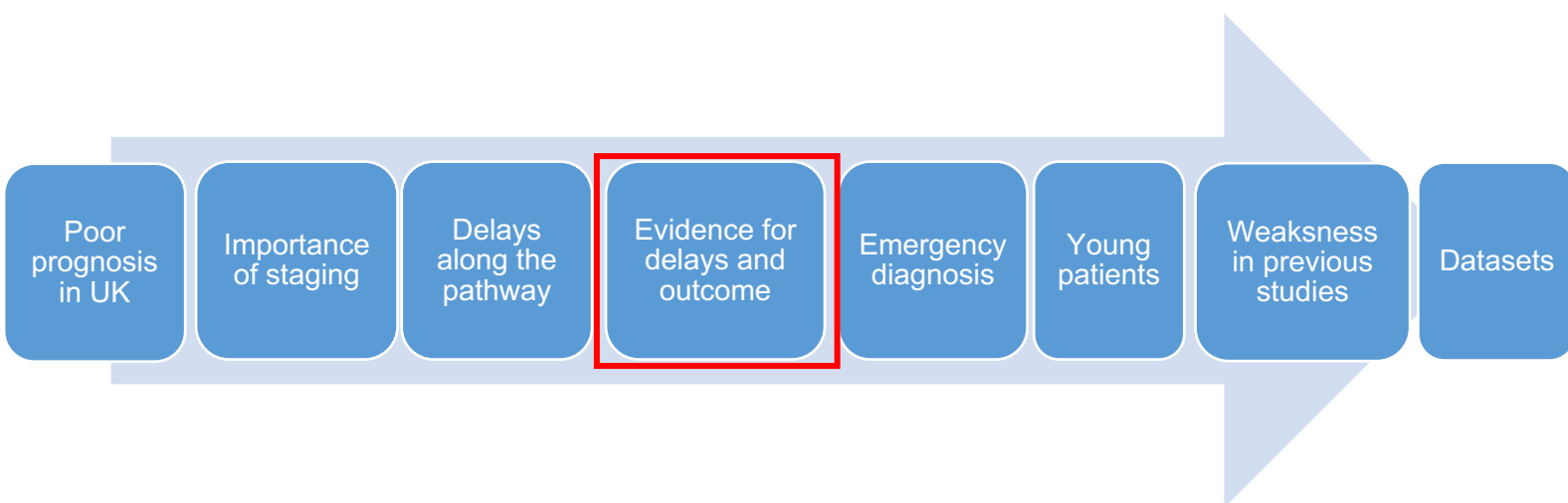
An initial misdiagnosis based on the presenting symptom can lead to inappropriate investigations³². A false negative result, despite the presence of symptoms has been shown to lead to a delay in referral.

The clinical acumen of the GP, and their personal experience of the patient, can affect outcome, although these factors are difficult to measure³⁶. However patients from practices with a higher referral rate than average have a better outcome and reduced stage at diagnosis for oesophagogastric cancer⁵⁴. Consultations with different doctors in primary care is also associated with a delay in referral⁵⁵.

Lack of abdominal or rectal examination can lead to a missed diagnosis, thereby delaying referral³². Early presentation by the patients following the onset of symptoms can also lead to a delay, as the duration of symptoms as stated in the 2005 NICE referral guidelines have not been met. This is no longer a criterion in the latest guidelines. Practices in rural locations or increased distance from a specialist centre are also likely to lead to a delay in referral⁵⁵.

2.3.1.2.3 Patient factors

Overall there was no difference in the referral interval between genders⁵⁶. Patients who previously frequently visited their GP for reasons not related to cancer, were more likely to experience a delay in referral³³. Younger patients were more likely to experience a delay³², usually because the presenting symptom was assumed to be related to a benign condition, or the age did not fulfil the referral criteria. Patients with higher co-morbidities were likely to be referred earlier⁵⁷, as were patients from a higher socio-economic class⁵⁸. There is no clear evidence that ethnicity makes a significant difference to the speed of referral^{31,32}



2.4 Evidence behind an association between delays and prognosis

In this section I describe the evidence regarding delays and survival or stage, stratified according to the various intervals in the diagnostic pathway. Studies which only investigated the interval between first symptom and presentation to primary care were excluded. Three extensive systematic reviews have been published, with the latest in 2015, to determine whether an association exists, two of which attempted to answer the question by using a meta-analysis. I describe these studies, as well as highlighting particular studies that were included. This is followed by an explanation of the limitations of previous studies and how I plan to overcome these issues in my thesis.

2.4.1 First symptom to diagnosis or treatment

Ramos et al.⁵⁹ carried out a systematic review to investigate whether delays along the whole pathway from first symptom to diagnosis or treatment was associated with a worse survival. The majority of studies found were based on hospital data, with some studies only including patients who underwent a resection. There were mixed results leading to the authors concluding there was not enough evidence to suggest an association between delays and outcome. In a second systematic review⁶⁰ by the same authors, in 26 of 37 studies no statistically significant association between delays and stage was seen, in four a longer

diagnostic interval was associated with a less advanced stage while in seven a shorter delay was associated with a less advanced stage. The meta-analysis of 13 studies failed to show an association, although colon cancers tended towards a less advanced disease with longer delays, while in rectal cancers a shorter delay was associated with less advanced disease. In a study published after these systematic reviews using retrospective analysis of patient notes, the diagnostic interval was longer for stage A disease than stage D disease⁶¹.

In a study by Thompson et al.⁶², consecutive patients who were diagnosed with bowel cancer in Wessex between 1991 to 1994, demonstrated no association between first symptom and treatment delays of over 6 months and the stage of disease. Duration of symptoms were obtained from both primary and secondary medical notes.

2.4.2 Initial presentation to referral

The type of symptom at presentation has also been the focus of investigation in terms of stage at diagnosis (table 4). Stapley et al.⁶³, by using an ordinal regression model with missing stage considered after stage IV, found rectal bleeding in colorectal cancer patients was associated with an earlier stage at diagnosis (OR 0.50 95% CI 0.31 – 0.79 p = 0.003), while mild anaemia was associated with a later stage at diagnosis (OR 2.2 95% CI 1.2 – 4.3 p = 0.021).

Table 4 - Stage according to presenting symptom⁶⁴

First symptom	n (%)	A	B	C	D	Not known	P-value*
Rectal bleeding	97 (28)	8 (8)	24 (25)	31 (32)	27 (28)	7 (7)	0.003
Abdominal pain	87 (24)	6 (7)	7 (9)	35 (43)	23 (28)	11 (13)	0.59
Diarrhoea	76 (22)	4 (5)	7 (9)	32 (42)	22 (29)	11 (14)	0.5
Constipation	38 (11)	2 (5)	5 (13)	17 (45)	9 (24)	5 (13)	0.64
Loss of weight	26 (7)	0	2 (8)	14 (54)	6 (23)	4 (15)	0.98
Mild anaemia	37 (11)	3 (8)	3 (8)	11 (30)	10 (27)	10 (27)	0.021
Severe anaemia	43 (12)	4 (9)	3 (7)	18 (42)	14 (33)	4 (9)	0.73

*Ordinal regression

Mortality (table 5) according to the first symptom also reflected the stage at diagnosis⁶³: rectal bleeding was associated with a reduced Hazards ratio of death (HR 0.57 95% CI 0.41 – 0.81 p < 0.001), while mild anaemia (HR 1.8 95% CI 1.1 – 2.7 p = 0.01) was associated with a worse prognosis.

Table 5 - Mortality according to presenting symptom to primary care⁶³

First symptom	Univariable		Multivariable ^a	
	Proportional hazard (CI)	P-value	Proportional hazard (CI)	P-value
Rectal bleeding	0.56 (0.41, 0.79)	0.001	0.57 (0.41, 0.81)	0.001
Abdominal pain	1.3 (0.93, 1.7)	0.14	1.3 (0.95, 1.8)	0.096
Diarrhoea	0.93 (0.67, 1.3)	0.7	1.1 (0.77, 1.5)	0.63
Constipation	1.3 (0.84, 1.9)	0.27	1.1 (0.79, 1.8)	0.6
Loss of weight	1.5 (0.94, 2.4)	0.092	1.5 (0.91, 2.3)	0.11
Mild anaemia	1.5 (0.98, 2.3)	0.064	1.8 (1.1, 2.7)	0.01
Severe anaemia	0.98 (0.65, 1.5)	0.93	0.79 (0.51, 1.2)	0.3

^aOther variables in model: age, emergency admission status, sex, and site of tumour.

Deng et al.⁶⁵ also found similar results in an analysis of patients diagnosed in China. Patients with rectal bleeding (OR 2.90 95% CI 1.29 – 6.54 p = 0.010), or the absence of a change in bowel habit (OR 0.37 95% CI 0.15 – 0.92 p = 0.033), were more likely to be diagnosed at an earlier stage (I or II).

2.4.3 Referral to diagnosis

There are consistent findings that the urgent referral pathway has shortened the hospital and total diagnostic interval, although the difference in survival between urgent and non-urgent routes has been inconclusive due to the heterogeneity of the studies. In a study published shortly after the introduction of the 2000 guidelines, Eccersley et al.⁶⁶ found no difference in stage at diagnosis between patients referred via the urgent and non-urgent pathway, which was confirmed by Flashman et al.⁶⁷ in 2004. Long term survival was not considered in either study. For 217 colorectal cancer patients diagnosed in one hospital trust between 2000 – 01 in the UK, no difference in survival or stage was noted between the routes⁶⁸. In a later study by Moller et al.⁶⁹ including patients with any cancers registered in the Exeter Cancer registry (not just colorectal) and linked to the national Cancer Waiting Database, found patients from GP practices who had a higher propensity to refer via the urgent TWW pathway demonstrated a better prognosis in terms of any cause of death compared with lower use of this referral pathway. This indicates the practices own referral

threshold, rather than the pathway itself, was associated with a better prognosis. Maclean et al.⁷⁰ noted in a large study containing 27 079 colorectal cancer patients diagnosed in England in 2012, that after adjustment for age, sex and deprivation, patients from GP practices with a higher than average endoscopy investigation rate were associated with a lower percentage of advanced cancer. Again like the previous study, this reflects the practice threshold, rather than the referral guidelines themselves. In a later study by Currie et al.⁷¹, only patients with a resection following rectal cancer were considered. Again no difference in stage or survival was noted, although those referred via the TWW were seen and diagnosed quicker. Ramsay et al.⁷² analysed 369 primary care records for patients diagnosed in the Aberdeen Royal Infirmary colorectal MDT. This was one of a few that found Dukes stage was less advanced in the routine referral group. However stage was only analysed in the patients who underwent a resection, thereby excluding those with Stage D. The maximum wait of two weeks before being seen in secondary care seems arbitrary, with no studies to back the theory that an urgent referral is associated with a better outcome.

In a study by NCIN using the national cancer registry data from 2012, 19% of patients referred via a non-urgent route were diagnosed with stage I compared to 14% in the TWW route⁷³. TWW patients were also significantly more likely to be diagnosed with Stage III cancer 32% vs 26%, but less likely to have missing stage (6% vs 12 %).

Wattacheril et al.⁷⁴ found a significantly shorter interval between referral to diagnosis in patient with more advanced stages than in those with less advanced stage. After adjusting for stage, there was no association between delays in diagnosis and survival.

2.4.4 Initial presentation to diagnosis

A study based on patients over the age of 66 identified from the SEER database⁷⁵, found an overall diagnostic delay of more than 8 months for colon cancer patients was associated with a worse all-cause mortality (HR 1.31 CI 1.08 – 1.58) but not cancer specific mortality compared to patients diagnosed between 14 - 59 days. Short diagnostic intervals of less than a week were associated with a worse all-cause mortality, which was no longer significant after adjustment for tumour stage and patient demographics. No associations were found for rectal cancer. Similarly, a study based on electronic primary care records in

Canada⁷⁶, which took into account patient demographics and tumour site (but not stage), found no association between delays in diagnosis (represented as quartiles) and worse stage or outcome. Emergency presentations were included. Colon and rectal cancers were analysed together. Data from patient interviews in China also found no association between delays and stage at diagnosis for colon or rectal cancer⁶⁵.

Stapley et al.⁷⁷ analysis of primary care records found no significant difference between longer duration of the diagnostic interval and later stage. For this study, the interval was split according to 1 – 30, 31 – 90, 91 – 180 and 181 – 365 days.

Torring et al.⁷⁸ analysed primary care data from Denmark, collected through GP questionnaires. The latter were classed as vague or alarm symptoms as conveyed by the GP at the time of presentation, rather than based on referral guidelines. Patients were identified from hospital inpatient and outpatient databases and verified against the cancer registry. Charlson score, emergency presentation, gender and age were included as covariates. The outcome measure was 3-year mortality. The diagnostic interval was analysed as both categorical and continuous. In the former model the interval was split into short (less than four weeks), medium (5 – 11 weeks) and long (12 weeks or more). Patients with vague symptoms at presentation were more likely to experience long interval to diagnosis. Patient presenting with alarm symptoms had an initially high hazards ratio of death in the first 5 weeks which diminished thereafter. This trend was not noticed with vague symptoms at presentation. Stage was not considered a confounder. This study was the first to demonstrate a non-linear relationship between delays in outcome, with a U-shaped curve for alarm symptoms.

In a follow up paper⁷⁹, the same authors analysed the data from three different sources – patient records, GP questionnaire and patient survey. The former was the only source with the date of diagnosis confirmed by histology, while the remainder relied on patient notes and recall respectively. The authors found the GP based study (as per the study described above) had the only positive findings of early and late intervals being associated with a worse 5 year survival. Again, a U-shaped association was found. There was no adjustment for co-morbidities, symptoms or tumour stage.

The same authors amalgamated data from six colorectal cancer cohort studies to demonstrate increasing and subsequently decreasing odds of advanced CRC with longer interval from presentation to diagnosis⁸⁰. For the interval between diagnosis and treatment, a reverse effect was seen - decreasing and subsequently increasing odds of advanced CRC with longer secondary care intervals. These findings were not seen when each of the cohorts were analysed separately.

In a study using CPRD⁸¹, the diagnostic interval was defined as the first presentation in the year leading up to cancer diagnosis, with the end of the pathway as the diagnosis date as per the cancer registry, stratified to <1 month, 1 – 2 months, 3 – 6 months and over 6 months. Patients diagnosed through an emergency were excluded. The authors included patient demographics, tumour stage and differentiation and radiotherapy or chemotherapy as potential confounders to survival. The latter was presented as 5 year, conditional 5-year and excess hazards ratio (EHR). Alert and non-alert symptoms were analysed separately, with colon and rectal cancers considered together. This study demonstrated only 36% of colorectal cancer patients presented with 'NICE-qualifying' symptoms, with longer diagnostic intervals for patients presenting with 'non-alert' symptoms. Colorectal cancer patients who presented with NICE-qualifying symptoms had a 14.7 percentage points (95% CI: 13.9–15.4) higher five-year relative survival than patients who presented with non-alert symptoms (RS alert: 57.0%; 95% CI: 54.1–59.8% vs RS non-alert: 42.3%; 95% CI: 40.2–44.4%). There was no association found between diagnostic interval and EHR. Patients presenting with 'NICE-qualifying' symptoms, a higher mortality for both shorter and longer diagnostic intervals was noted, but did not reach significance. Similarly although there was a trend that a shorter diagnostic interval was associated with increased mortality following presentation with non-alert symptoms, this did not reach significance.

In a study based on routine primary care data from Northern Scotland, colorectal cancers diagnosed and confirmed in the cancer registry between 1997 – 98 were included⁸². The interval was defined from the first presentation to definitive treatment, or diagnosis if no treatment was given. The interval was considered as a continuous variable, with outcome measured as survival or stage of cancer. The models included symptom at presentation,

stage, grade of differentiation, age, Charlson score, type of admission, gender, smoking status and Carstairs deprivation index. There was a worse prognosis for colorectal cancer patients who had a short diagnostic interval, which was no longer apparent in the fully adjusted model. Separate colon and rectal analysis failed to show a difference. The authors demonstrated a non-linear relationship between delays and outcome, although the findings did not reach significance. Rectal bleeding at presentation independently reduced the hazards ratio of death while anorexia increased the ratio. There was no difference in findings whether survival was measured from first presentation or diagnosis.

2.4.5 Onset of symptoms to diagnosis

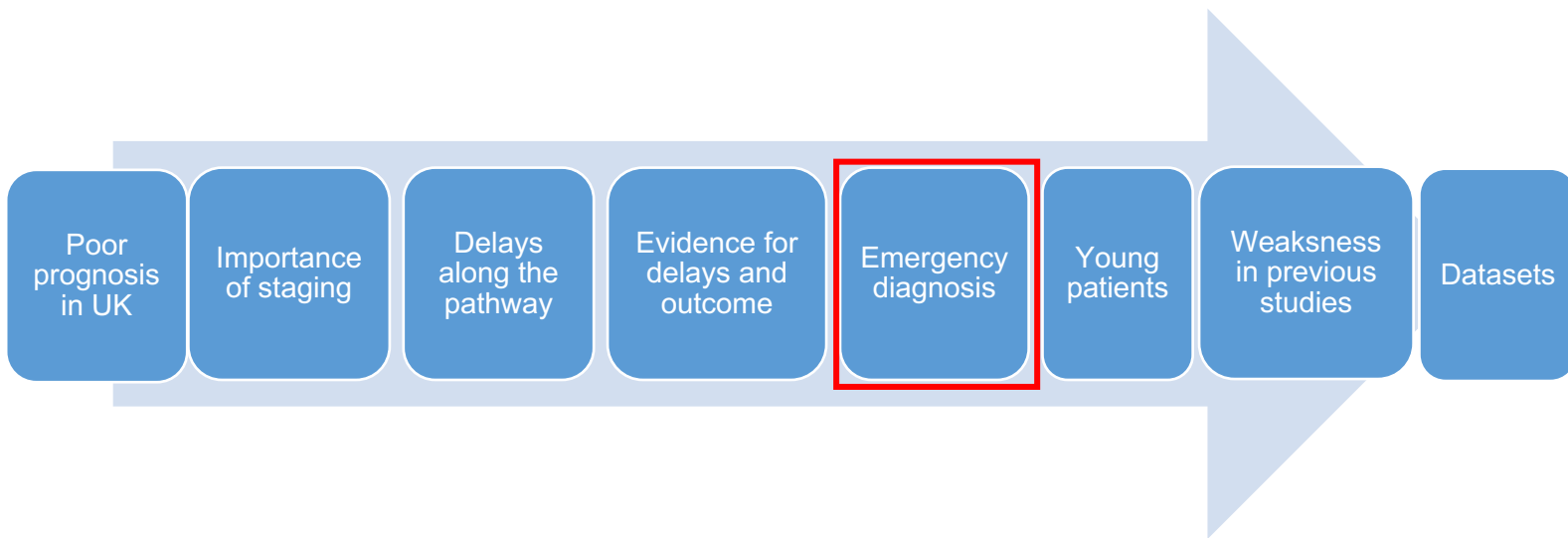
Walter et al.³⁶ found early stage colorectal cancer was associated with a longer total diagnostic interval (from start of symptoms to diagnosis) for all patients (early stage median interval 157 (93 – 243) vs late stage 99 (65 – 198)). Early stage cancer was associated with a longer total diagnostic interval even after stratifying the symptoms according to bleeding per rectum, change in bowel habit, abdominal pain and weight loss. However test for significant difference was only carried out for all symptoms, presumably because of the low number of patients with a single symptom.

2.4.6 Diagnosis to treatment

In a study by Pruitt et al.⁷⁵ mentioned previously, treatment delays of less than a week were associated with worse overall survival for both colon and rectal cancer compared to delays of 1 – 2 weeks. In these models stage was included as a co-variate⁷⁵. This paradoxical finding was confirmed by Gort et al.⁸³. who demonstrated patients undergoing a resection within seven weeks from diagnosis demonstrated a significantly lower odds ratio for disease free (OR 0.69 CI 0.51 – 0.94) and relative survival (0.66 CI 0.44 – 0.99). Terhaar et al.⁸⁴ also found late stage disease had a shorter interval from diagnosis to treatment (4.9 vs 3.6 weeks). However in two separate studies, rectal cancer patients who experienced a delay of over 31 days had a worse overall survival⁸⁵, while colon cancer patients who experienced a diagnosis to treatment interval of over more than 30 days had an increased odds ratio for a late stage diagnosis (III or IV). In contrary to these studies, Roland et al.⁸⁶ found no association with longer intervals between diagnosis and treatment.

2.4.7 Summary of associations between delays and outcome

There are inconsistencies as to whether delays along the diagnostic pathway, and in particular in primary care, are associated with a worse survival for colorectal cancer patients who are diagnosed as a non-emergency. This is due to a lack of consistency in definitions of the intervals along the patient pathway, and insufficient inclusion of confounders which may explain the findings. By using CPRD I am aiming to include sufficient detail regarding patients symptoms, demographics and tumour location to determine associations between delays and outcome.



2.5 Emergency diagnosis as a marker of diagnostic delays

For this section I will be testing the following hypothesis:

Patients experience missed opportunities in primary care to prevent an emergency diagnosis

The research questions are:

- Which patients are most likely to be diagnosed as an emergency?
- Which symptoms are most likely to lead to an emergency diagnosis after accounting for patient demographics?
- Is there are difference in GP usage between emergency and non-emergency patients in the four years leading up to diagnosis?
- Could emergency patients have been identified earlier to prevent an emergency diagnosis?
- Are emergency patients with the same symptoms as non-emergency patients diagnosed at a similar tumour stage?

In the following chapter I discuss the current knowledge regarding factors associated with an emergency diagnosis for colorectal cancer patients and the difference in prognosis compared with the non-emergency route.

2.5.1 Why is this important?

Despite the introduction of NICE guidelines, the percentage of patients diagnosed as an emergency has not improved significantly⁸⁷. Emergency diagnoses, for all cancers, are associated with a later stage and worse overall prognosis⁸⁸. In addition patient satisfaction is lower following such an emergency diagnosis⁸⁹. Even if a patient were to present with the same disease stage, emergency patients are less likely to receive chemotherapy and curative surgery⁹⁰. Overall there is limited evidence of how emergency diagnosis can be avoided. Recent reviews have attempted to classify reasons into modifiable and non-modifiable risk factors^{91,92}.

Non-modifiable factors are mostly based around the characteristics of the tumour. Patients with aggressive tumours that manifest as bowel obstruction or perforation with no or little warning are difficult to identify^{88,93}. Although these patients may have a preference of primary over secondary care, the acute and severe nature of the symptoms simply do not allow the time for diagnosis through a primary care route⁸⁸. Such tumours may have been identified through a screening programme before presenting at a late stage, although it would be necessary to widen the screening criteria to the youngest and oldest, for whom tumours are likely to be more aggressive⁹⁴.

Modifiable factors include both patient and practitioner factors^{93,94}. There is a cohort of patients who prefer to seek medical attention in secondary care due to the relatively quicker process of consulting with a doctor, some of whom would be perceived as a 'specialist' in cancer as opposed their own general practitioner. On the other hand a patient may wait, in fear of the diagnosis, until the symptoms become severe enough to warrant an emergency presentation. Such factors are mentioned in the description below but are not the focus of my thesis.

Modifiable factors related to GPs, form the basis of this chapter. These include the symptoms and patient demographics that are most likely to lead to an emergency diagnosis, despite prior presentation to primary care. A better understanding of why such patients 'slip through the net' could improve their identification and so reduce the emergency diagnostic rate.

2.5.2 Definition of an emergency diagnosis

The Oxford Dictionary defines an emergency as

A serious, unexpected, and often dangerous situation requiring immediate action

In the context of clinical studies, the term emergency has been used to describe either the perception of the patient or primary care physician that the symptom or situation is sufficiently 'serious, unexpected or dangerous', or simply describes the attendance to the emergency department, irrespective of the symptoms.

2.5.2.1 Based on patients' symptoms or procedure

The source for such studies includes analysis of patients' medical records, patient interviews, audits or large national datasets. The heterogeneity in definitions and methodology has led to differences in the percentage of diagnoses defined as an emergency, and the associated factors. For example Pruitt et al.⁹⁵ analysed the SEER database to identify patients who had an emergency surgery. A similar definition was used by Cleary et al.⁹¹, who sourced patients from a region in the UK (Exeter) and analysed their primary care records. Gunnarsson et al.⁹⁶ further refined the definition of an emergency presentation as those that were admitted via Accident and Emergency, had an operation within three days of admission and were found to have obstruction, perforation or bleeding at the time of the procedure. These studies fail to identify the two ends of the spectrum - cases where the GP or the patient deemed an emergency consultation was necessary, but did not require an operation, and those too unwell to undergo a procedure.

2.5.2.2 Emergency diagnosis defined as attendance to Accident and Emergency

The alternative method is to define an emergency diagnosis as all patients who attend Accident and Emergency, irrespective of the treatment outcome. Recent studies from the UK have used the algorithm by Elliss-Brookes et al.²⁹ in the 'Routes to diagnosis' study, which separates the diagnostic route into eight categories based on data from national linked datasets. The data provided by CPRD for my thesis provides the diagnostic route based on this algorithm, and so separates emergency and non-emergency patients. A description of this algorithm follows.

2.5.2.2.1 Routes of Diagnosis algorithm

For the 'Routes to diagnosis' study²⁹, patients were extracted from the National Cancer Data Repository, and linked via the unique NHS number to Inpatient and Outpatient HES, the National Cancer Waiting Times Monitoring data (CWT) and NHS Bowel Cancer Screening Programme (NBCSP).

The algorithm initially identifies the end-point of the route as the episode in HES that lead most immediately to diagnosis. If the date of diagnosis in the cancer registry matches an inpatient or outpatient HES activity, then the inpatient episode takes priority as the end-point. If no activity is found on the date of cancer diagnosis, the 28 days leading up to diagnosis is analysed, with inpatient taking preference over outpatient, and the most recent episode if there are multiple episodes. If there is no HES activity within 28 days, the process is repeated for data up to six months before the diagnosis date, with the end-point dependent on the nearest HES episode to the diagnosis date, irrespective whether it is an inpatient or outpatient episode. If there is no HES activity in the six months before diagnosis, then the route is labelled as 'Unknown' or 'Death Certificate Only', the latter based on information from the individual cancer registries. Therefore the end-point can be defined as 'Outpatient', 'Inpatient', 'DCO' or 'unknown'. For the former two, the start point of the route is established.

A diagnosis is deemed an emergency if the inpatient end-point is an emergency admission, or if an emergency referral preceded the outpatient end-point. From an outpatient end-point, it is also possible to define the route as routine GP, other outpatient or TWW route dependent on the referral data from the Cancer Waiting Times. Patients with 'Unknown' routes have no data in inpatient or outpatient HES, Cancer Waiting Times or Screening. The eight routes, and their definitions are:

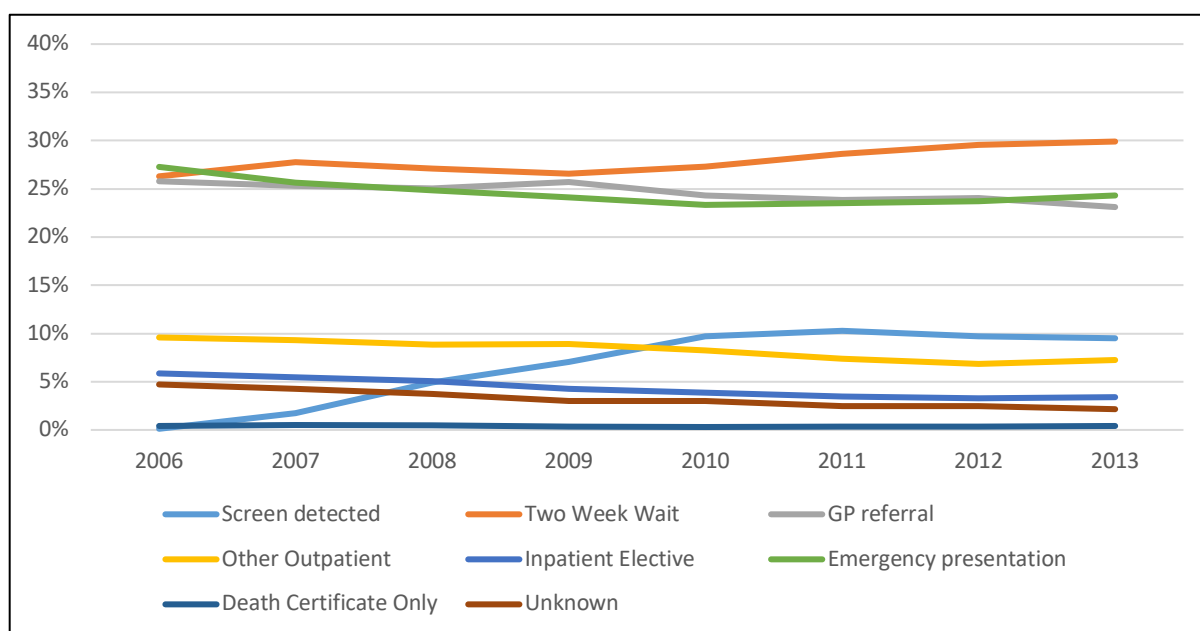
- *GP Referral*: includes routine and urgent referrals where the patient was not referred under the Two Week Wait referral route
- *Two Week Wait*: urgent GP referrals with a suspicion of cancer
- *Emergency Presentation*: an emergency route via A&E, emergency GP referral, emergency transfer, emergency admission or attendance

- *Other Outpatient*: an elective route starting with an outpatient appointment that is either a self-referral, consultant to consultant referral, other or unknown referral (these referrals would not include patients originally referred under the Two Week Wait referral route)
- *Screen Detected*: flagged by the cancer registry as detected via the screening programme
- *Inpatient Elective*: where no earlier information can be found prior to admission from a waiting list, booked or planned.
- *DCO*: diagnosis by death certificate only
- *Unknown*: no relevant data available from IP or OP HES or from NCWT or screening.

Studies that have used this algorithm have simplified the routes into emergency and non – emergency, with the latter including GP referral, Two Week Wait, Inpatient Elective and Other Outpatient. In one study, Unknown was also deemed a non-emergency route⁹⁷.

Through this algorithm, 22% of all colorectal cancers between 2006 – 2013 in the UK were diagnosed as an emergency, with a modest fall from 27% in 2006 to 24% in 2013 (figure 5)⁷³. During this period the introduction of screening, and the increase in two week wait diagnoses is apparent, leading to reductions in the Inpatient Elective, Unknown and Other Outpatient routes.

Figure 5 - Percentage of colorectal cancer diagnoses by presentation route from 2006 – 2013. Data source NCIN⁶⁸



The emergency route is further defined as:

- *Accident & Emergency:* Admitted from the A&E department
- *GP referral:* Admitted to inpatients from a GP as an emergency referral
- *Inpatient Emergency:* Admitted to inpatients from the Bed Bureau as an emergency referral, via Mental Health Crisis Resolution Team or admitted to inpatients from a consultant outpatient clinic
- *Outpatient Emergency:* Referred from A&E attendance to outpatients under the care of referring consultant, following an emergency inpatient admission or from A&E attendance to outpatients under the care of different consultant

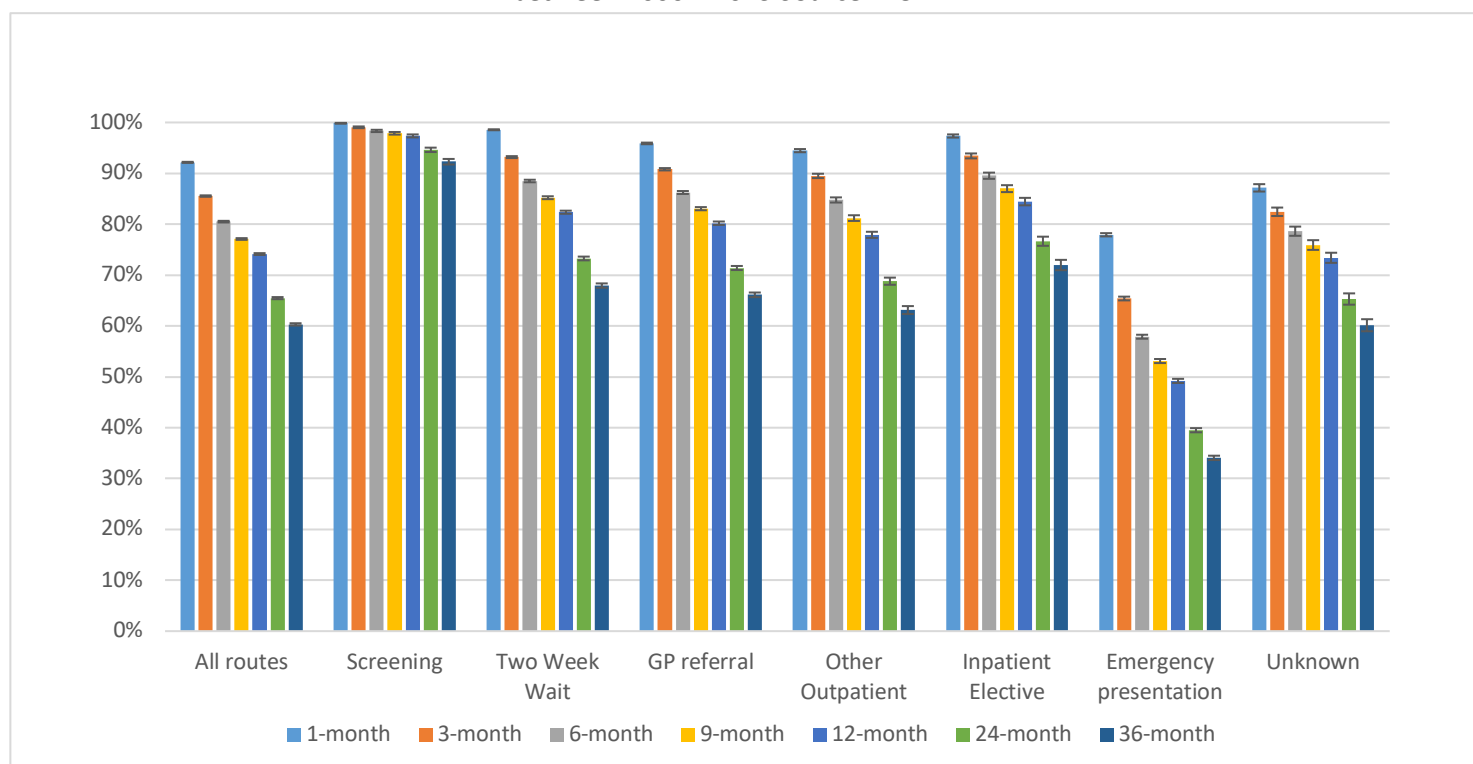
In 2006, 57% of patients presented directly to A&E, 34% referred as an emergency by their GP, 4% as an Inpatient Emergency and 6% as an Outpatient Emergency. By 2014 those presenting to A&E directly had increased to 68%, while 21% were referred by the GP, indicating more patients are bypassing their GP before presenting as an emergency⁹⁸.

2.5.3 Survival following an emergency diagnosis

The relative survival at one year for emergency patients is 50% (95% CI 49 – 51), compared with 82% (95% CI 82 – 83) for the two-week-wait, 82% (82 – 83) for GP referral, 80% (95% CI

79 – 81) for other outpatient, 84% (95%CI 83 – 85) for inpatient elective, 73% (95% CI 72 – 74) for Unknown and 98% (95% CI 97 - 98) for screen detected. Survival for emergency cancer patients drops from 78% at one month to 34% (56.4% relative fall) at three years, as opposed to a decrease from 99% to 68% (31.3% relative fall) for TWW patients. However after adjusting for age, gender, co-morbidities, economic deprivation and stage¹⁶, the excess mortality rate ratios (EMRR) for patients with colorectal cancer diagnosed as an emergency versus electively were found to be 5.9 (95%CI 5.6–6.2), 2.4 (95% CI 2.3–2.6), 1.8 (95% CI 1.7–1.9) and 1.7 (95% CI 1.6–1.8) at 1, 3, 6 and 12 months after diagnosis, respectively (figure 6). This indicates difference in long term survival narrows the longer the patient survives after diagnosis. Emergency rectal cancer patients of similar stage are less likely to receive a resection as non-emergency patients⁹².

Figure 6 - Relative survival by presentation route for colorectal cancer patients diagnosed between 2006 – 2013 Source: NCIN⁹⁹



The stage at diagnosis for non-emergency colorectal cancer patients (excluding Unknown and DCO) in 2012-2014 was 15.5%, 23.6%, 26.9%, 20.4% and 13.5% for stage A, B, C, D and stage unknown respectively⁹⁹. In comparison emergency patients were diagnosed at a later stage – 4.7%, 19.0%, 19.7%, 31.5% and 25.1% respectively⁹⁹.

The later stage at diagnosis for emergency colorectal cancer patients, and the relatively poor survival for this group, demonstrates further work needs to be done on a national level to reduce the proportion of patients who are diagnosed as an emergency. The relatively static emergency diagnosis rate between 2006 and 2013⁹⁹, despite the introduction of screening and new NICE referral guidelines, indicates a lack of understanding of the reasons that lead to an emergency diagnosis. Below I describe the current knowledge surrounding possible causes.

2.5.4 Factors associated with an emergency diagnosis

2.5.4.1 Patient demographics

Patients at opposite ends of the spectrum in age are more likely to be diagnosed as an emergency. 29% of those under 50 years of age were diagnosed as an emergency in the UK between 2006 – 2013. This compares to 20%, 17%, 22%, 29% and 42% of those aged 50 – 59, 60 – 69, 70 – 79, 80 – 84 and 85 and over, respectively. For those aged under 50, the proportion of emergency diagnoses has risen from 28% in 2006 to 34% in 2013, while the emergency rate has fallen for all other age groups⁹⁹.

In a multivariate model for national data from the UK, the odds ratio for an emergency diagnosis was significantly higher for females, Dukes B, C or D (compared with Dukes A), age 0 – 24, 65– 84, 85+ (compared with 25 – 64), Charlson co-morbidity score of 1, 2, 3+ (compared with score 0), and all levels of Index of Multiple Deprivation (compared with IMD = 1)¹⁰⁰. Similar findings have been described following analysis of the SEER database, which also found patients from Asian (OR 1.16 95% CI 1.08 – 1.24) or Black (OR 1.12 95% CI 1.05 – 1.18) backgrounds were more likely to be diagnosed as an emergency compared with White patients⁹⁵.

Abel et al¹⁰¹. analysed the survey data for 4647 emergency cancer patients from the 2010 English Cancer Patient Experience Survey, which was linked to the Routes of Diagnosis data as per the Elliss-Brookes et al.²⁹ algorithm. This included various cancers, not only colorectal. Patients diagnosed as an emergency were defined as either those with no prior GP usage or those who had seen their GP 3 or more times before diagnosis. The aim of this study was to compare the GP usage of various cancers compared to colorectal cancer, as well as the

patient demographics most likely to lead increased consultations before an emergency diagnosis. They found a third of emergency patients had no previous contact with their GP. Older and more deprived patients were more likely to have not consulted their GP before an emergency diagnosis. The young or those from an ethnic minority were more likely to have seen their GP multiple times. The GP usage was based on questionnaires completed by patients rather than review of patient notes, potentially introducing recall bias. In addition, the symptoms related to colorectal cancer and emergency diagnoses was not considered.

2.5.4.2 Practice factors

In a study using CPRD by Tsang et al.¹⁰², the relative risk of an emergency diagnosis increased the longer a patient was known to a practice (time at practice split into tertiles: moderate (RR 1.59 (1.31 – 1.91) $p < 0.0001$) or high (RR 2.06 (1.72 – 2.46) $p < 0.0001$)). As there was no association between the continuity of care received by the patient (related to the amount of consultations by each GP in a practice), these findings indicate patients familiar to the practice are more likely to be diagnosed as an emergency, irrespective as to whether they are consulted by same GP each time. In another study the same authors⁵⁵ found practices with a higher Quality and Outcomes Framework (QOF) score reduced the odds ratio of an emergency admission (OR 0.94 95% CI 0.91 – 0.97) as did practices that provided an appointment within 48 hours (OR 0.85 95% CI 0.79 – 0.92). However patients registered in practices where only some (OR 1.04 95% CI 1.02 – 1.06 $p = 0.001$) or none (OR 1.08 95% CI 1.04 – 1.11 $p < 0.0001$) of the GPs qualified from the UK, or those in the most deprived areas, were more likely to be diagnosed as an emergency. However these findings were based on a combination of cancers, not just colorectal.

Patients with preventable hospitalisations in the year leading up to colorectal cancer diagnosis were more likely to be diagnosed as an emergency in two US based studies using the SEER⁹⁵ and a state-wide database¹⁰³ indicating a preference for the use of secondary rather than primary care. Colorectal cancer patients with regular access to primary care were also less likely to present as an emergency in a study based in Canada¹⁰⁴.

Oesophagogastric cancer patients from practices with a higher gastroscopy rate per capita practice population are less likely to be diagnosed as an emergency, more likely to undergo

a resection and demonstrate better prognosis, suggesting a lower threshold for diagnostic procedures improves patient outcome⁵⁴. Investigations, such as colonoscopies, in the year leading up to a colorectal cancer diagnosis were also less likely to lead to an emergency diagnosis, again re-enforcing the concept that an improved awareness of the potential for symptoms to be cancer related can reduce the rate of emergency diagnoses^{96,104}.

However studies have also shown the incorrect investigations, such as the use of an abdominal ultrasound scan for patients with underlying colorectal cancer is associated with an emergency diagnosis. Some patients may present as an emergency whilst awaiting the results of appropriate investigations or an outpatient specialist appointment. In a small study, 10 of 39 patients who were diagnosed as an emergency were awaiting for such an appointment⁴⁸.

2.5.4.3 Patient symptoms

Four studies have described the difference in symptoms between emergency and non-emergency patients. These studies are heterogeneous in their data source and analysis, with three based in the UK and one from Sweden. In all the studies the time leading up to diagnosis was analysed for differences in emergency and non-emergency symptoms rather than the presenting symptom.

Cleary et al.⁹¹ (table 6) identified colorectal cancer patients from an area within the UK aged 40 years and above, diagnosed between 1998 and 2002 inclusive, and case matched them in a 5:1 ratio by age and gender to patients without cancer. Emergency cases were defined as those with an emergency operation for bowel obstruction or perforation. Primary care records were individually analysed for the presence of symptoms in the two years leading up to diagnosis. Any episodes in the 30 days before diagnosis were excluded, as these consultations may have directly led to the diagnosis, and so do not represent missed opportunities. Abdominal pain (OR 6.2 95% CI 2.8 – 14.4 p < 0.001), loss of weight (OR 3.4 95% CI 1.3 – 8.5 p = 0.01) and diarrhoea (OR 2.6 95% CI 1.2 – 5.7 p = 0.02) were positively associated with an emergency case. Rectal bleeding, constipation, tenderness of palpation, haemoglobin level less than 13 g/dl and a blood sugar level of over 10mmol/l were also associated with an emergency case on univariate analysis, but were not included in the multivariate model due to the relatively low numbers.

Renzi et al.¹⁰⁵ (table 6) identified colon or rectal cancer patients aged 25 years or older with a confirmed diagnosis in the cancer registry with linked data in CPRD. In the 25 – 60 months before a cancer diagnosis, only rectal bleeding was significantly more common in the non-emergency colon cancer group (3.9% vs 1.7% $p = 0.049$). No other symptom was more frequent in either non-emergency or emergency cases. However between 30 days to 12 months, emergency colon cancer patients were less likely to present with rectal bleeding, anaemia or change in bowel habit, while only rectal bleeding was less common in emergency patients for rectal cancers. These same symptoms remained significant after accounting for age, gender, previous GP consultations and socioeconomic deprivation. Emergency was defined according to the 'Routes to Diagnosis' algorithm. The symptoms at the initial presentation to the GP was not described.

Gunnarsson et al.¹⁰⁶ (table 6) presented data from Sweden for colorectal cancer patients with a registration in the regional cancer registry between 2006 – 2008. Primary care data for 12 months before the cancer diagnosis was analysed for relevant symptoms in an electronic database for the majority of patients, with paper records obtained for the remainder (number not available). The final month before diagnosis was included. Patients were defined as 'acute emergency' if the patients was admitted through Accident and Emergency, had surgery within three days of admission and obstruction, perforation or bleeding were confirmed during the operation. 'Subacute' cases were patients who were determined in the cancer registry as emergency, but did not fulfil all three criteria above. Constipation and abdominal pain was once again more common in the acute patients, while rectal bleeding and anaemia were less frequent. 34 (26.2%) of the acute and 17 (18.1%) of the subacute cases had no GP recording in the 12 months leading up to the cancer diagnosis. Multivariate analysis was not carried out. Once again simply the presence of the symptom, rather than the symptom at initial presentation, was considered.

Sheringham et al.⁹⁷ (table 6) extracted colorectal cancer patients from a regional cancer registry in England. They obtained the primary care records, and the inpatient and outpatient hospital records for all patients, although this was not HES based. They modified the 'Route to Diagnosis' algorithm to only consider episodes that were related to a

resection, colonoscopy or inpatient treatment by gastroenterology or surgery. They produced four routes: 1. Emergency presentation (including GP referrals), 2. GP urgent/2 week wait, 3. GP routine/unknown 4. Consultant, other or unknown. A screening identifier was not available to allow separation of this route. The latter three routes were amalgamated into the elective route and compared to the emergency route. Twelve months of data was analysed for relevant symptoms. Stage data was also included as early (Stage I and II) (28.0%), late (Stage III and IV) (39.8%) and missing stage (32.2%), which was collapsed into a binary variable by combining the unknown with late stage. As stated by the authors, the reason for the latter is that a higher proportion of cancer staging information has been found to be missing in patients that die rapidly after diagnosis (Downing et al, 2013). They found rectal bleeding reduced the odds of an emergency diagnosis (OR 0.18 95% CI 0.07 – 0.43 $p < 0.001$), although constipation (OR 1.85 95% CI 1.20 – 2.86 $p < 0.01$) and abdominal pain (OR 2.32 95% CI 1.46 – 3.69 $p < 0.0005$) were associated with an emergency diagnosis, as was a late/unknown stage at diagnosis (OR 1.52 95% CI 1.01 – 2.28 $p < 0.05$).

Table 6 - Percentage of patients with cancer related symptoms from four studies

	Cleary et al. ⁹¹		Renzi et al. ¹⁰⁵				Gunnarsson et al. ¹⁰⁶			Sheringham et al. ⁹⁷		
Data preceding diagnosis	30 days to 24 months		30 days to 12 months				12 months			12 months		
Definition of emergency	Emergency operations		Routes of diagnosis algorithm				Emergency operation within 3 days of admission			Modified Routes of diagnosis		
Patient source	Cancer Registry		Cancer Registry				Cancer registry, only those with an operation			Cancer registry		
GP data source	Retrospective analysis of primary care notes, matched controls		CPRD				Electronic database and patient notes			Retrospective analysis of primary care records		
Site	Colorectal		Colon		Rectal		Colorectal			Colorectal		
Emergency	Y	N	Y	N	Y	N	Acute	Subacute	N	Y	GP	Other
Number	62	310	361	668	87	490	130	94	508	228	536	179
Abdominal pain	39	7.1	24.4	25.5	12.6	8.8	45.4	43.6	35.2	21.9	16.2	10.6
Loss of weight	19	5.2	3.1	3.1	2.3	1.6	12.3	25.5	16.9	2.6 ^b	17.9	6.7
Diarrhoea	29	9.4	10.0	9.9	16.1	11.2	16.9	29.8	24.2	6.4	9.0	5.6
Rectal bleeding	11	3.9	3.6	12.6	9.2	23.7	4.6	10.5	21.1	2.6	17.9	6.7
Constipation	16	7.4	6.9	8.1	5.8	8.6	25.3	29.8	19.3	24.1	17.0	11.2
Anaemia	15 ^a	3.6	11.9	23.2	8.1	5.9	28.5 ^b	44.7	55.3	40.8	50.2	27.9
Change in bowel habit	-	-	3.3	6.7	8.1	12.2	-	-	-	2.2	2.6	4.5
Fatigue	-	-	3.3	4.9	1.2	2.5	-	-	-			

^aHaemoglobin level below 13 g/dl ^bIncludes fatigue

2.5.4.4 Summary of factors associated with an emergency diagnoses

The difference between patients diagnosed as an emergency and those through a non-emergency route has previously been investigated as described above, but not in a single study that includes the important confounders. These studies lack in patient numbers, define emergencies based on an operation, or fail to exclude the last 30 days before diagnosis. This period should not be included in attempting to identify missed opportunities, as it is possible GP consultation during the last 30 days directly led to the emergency diagnosis. By using CPRD, I am aiming to overcome the deficiencies in previous studies.



2.6 Colorectal cancer delays for young patients

For this chapter the hypothesis states:

Young patients with colorectal cancer experience missed opportunities and delays in primary care

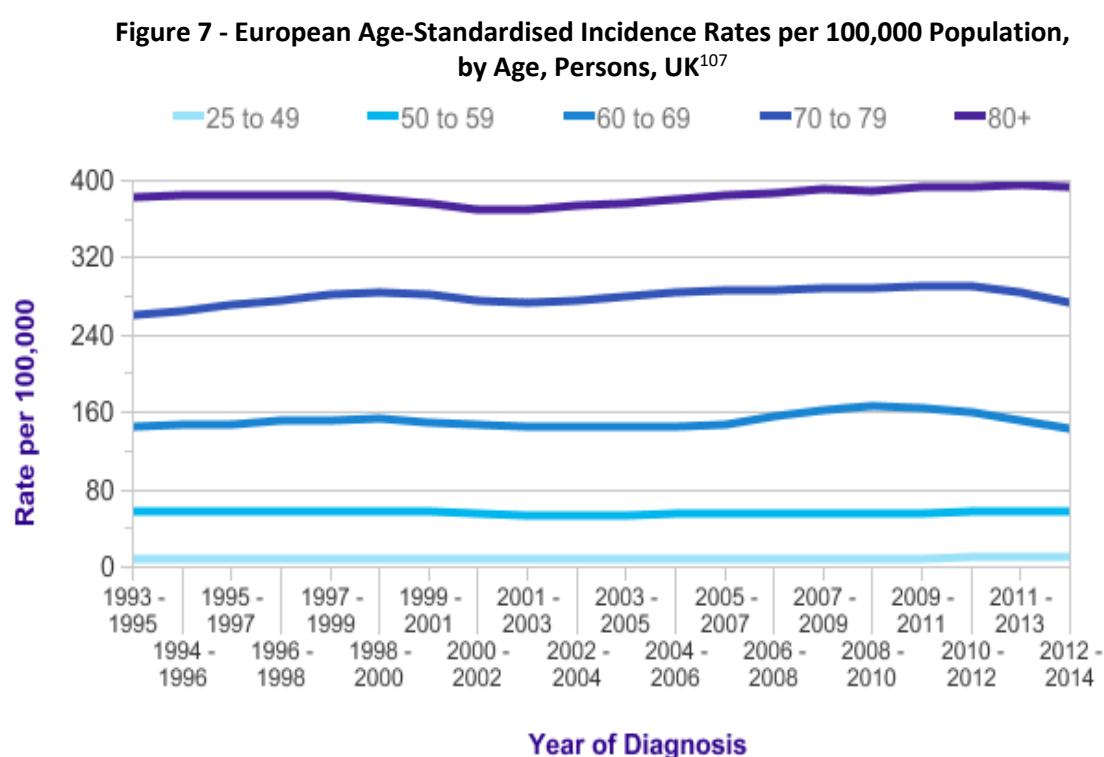
The research questions for this chapter are:

- Are there any differences in the presenting symptom between young patients, defined as under 50, and the older age groups?
- What is the difference in the primary, hospital and diagnostic interval between the age groups?
- Are young patients more likely to be diagnosed as an emergency even if they present with the same symptoms as older patients?
- Is there a difference in stage at diagnosis between young and older patients despite presentation with similar symptoms?

Young colorectal cancer patients, defined as those aged under 50, form the focus of this chapter. I initially describe the growing incidence of colorectal cancer in this age group, noted both in the UK and USA. Despite promotion of the awareness of colorectal cancer in primary care by NICE, colorectal cancer diagnosis in young patients is more likely to be diagnosed at a later stage. The evidence for this is described below, together with possible associations with delays in diagnosis in primary care.

2.6.1 Incidence of colorectal cancer in the young

Since the early 1990s, bowel cancer incidence rates have increased by less than 4% in the UK, with an increase of less than 2% in the last decade¹⁰⁷. Overall incidence rates are projected to fall by 11% in the UK between 2014 and 2035. Analysis according to age has shown the incidence has remained stable for those aged 50 – 59 and 60 - 69, but the European Age-standardised rate in patients under 50 has increased by 34% between 1993 – 1995 and 2012 – 2014¹⁰⁷(figure 7), although this represents a small absolute rise from 16 to 22 per 100 000 population. Data stratified according to subsite is not available.



The incidence of colorectal cancer in young patients in the US is also increasing. A recent publication from the Surveillance, Epidemiology, and End Results (SEER) database analysed trends across age groups between 1974 to 2013¹⁰⁸ for 490 035 cases. These were organised according to site; proximal colon (C18.0, C18.2- C18.4) distal colon (C18.5-C18.7) rectum (C19.9, C20.9). In the 1970's and 1980's, the incidence of colon cancer was increasing in all age groups over the age of 50, while decreasing in those under 50. However since the 1980's, the converse has occurred. Incidence in those aged 55 years and older has decreased, while increasing by 2.4% per year in adults age 20–29 years and by 1.0% per year in adults age 30– 39 years. In the mid-1990s, rates also began increasing in adults age 40 to

49 years (1.3% per year) and 50 to 54 years (0.5% per year). Increasing trends in adults younger than age 50 years were confined to tumours in the distal colon, with the exception of adults age 40 to 49 years, among whom rates are also increasing for proximal tumours. The lower incidence in the older age groups is reflected by the decline in distal site tumours. Rectal cancer incidence rates show a similar trend – an increase by 3.2% per year from 1974 to 2013 in adults age 20 to 29 years and since 1980 in adults age 30 to 39 years, and by 2.3% per year in adults age 40 to 49 years and 50 to 54 years. In contrast, rates generally declined throughout the entire 40-year study period in adults age 55 years and older. This trend was first noticed in 2003 by O’Connell et al., again using SEER data¹⁰⁹. You et al.¹¹⁰, who analysed data from the US National Cancer Database, containing 70% of all US incident cancers annually¹¹¹, noted CRC in young patients increased from 1998 to 2007, with an annual percent increase of 2.1% (95% CI, 1.1%-3.1%), whereas the incidence in older patients decreased by 2.5% (95% CI -3.0% to -2.0%). As in SEER, young-onset rectal cancer incidence (annual increase of 3.9%; 95% CI, 3.1%-4.7%) increased more rapidly than young-onset colon cancer (annual increase of 2.7%; 95% CI, 2.0%-3.3%).

2.6.2 Prognosis in young colorectal cancer patients

Outcome for young CRC patients can be compared to the older population by measuring survival from diagnosis to cancer specific mortality or using the stage at diagnosis as a surrogate marker. Some papers have been able to describe both.

Using national data, the mortality in the US seems to differ from the UK. ONS data, which uses European Age-Standardised Mortality Rates, demonstrates a decrease in mortality for all age groups, and in particular a fall of 55% for those aged 25 – 49 between 1971 – 2014 (5.5 deaths per 100 100 in 1971 – 1973; 4.6 in 1980 – 1982; 3.7 in 1990 – 1992; 2.6 in 2000 – 2002; 2.5 in 2012 – 2014) (Figure 8)¹⁰⁷. In the SEER data¹⁰⁸, which also uses age-adjusted rates, the mortality remained steady for patients aged under 50 from 1992 to 2004, but increased by 1.4% annually since 2004 (figure 9). Mortality in all other age groups decreased significantly. The difference could be due to a difference in the start of data coverage. If the ONS data is similarly restricted from 1992, the difference in mortality is noticeably less (1992 – 1994 3.4 deaths per 100 100 in ONS compared with 2.8 in SEER). Data from 1972 is not available on the SEER website.

Figure 8 - European Age-Standardised Mortality Rates per 100,000 Population, by Age, Persons, UK for patient diagnosed with colorectal cancer¹⁰⁷

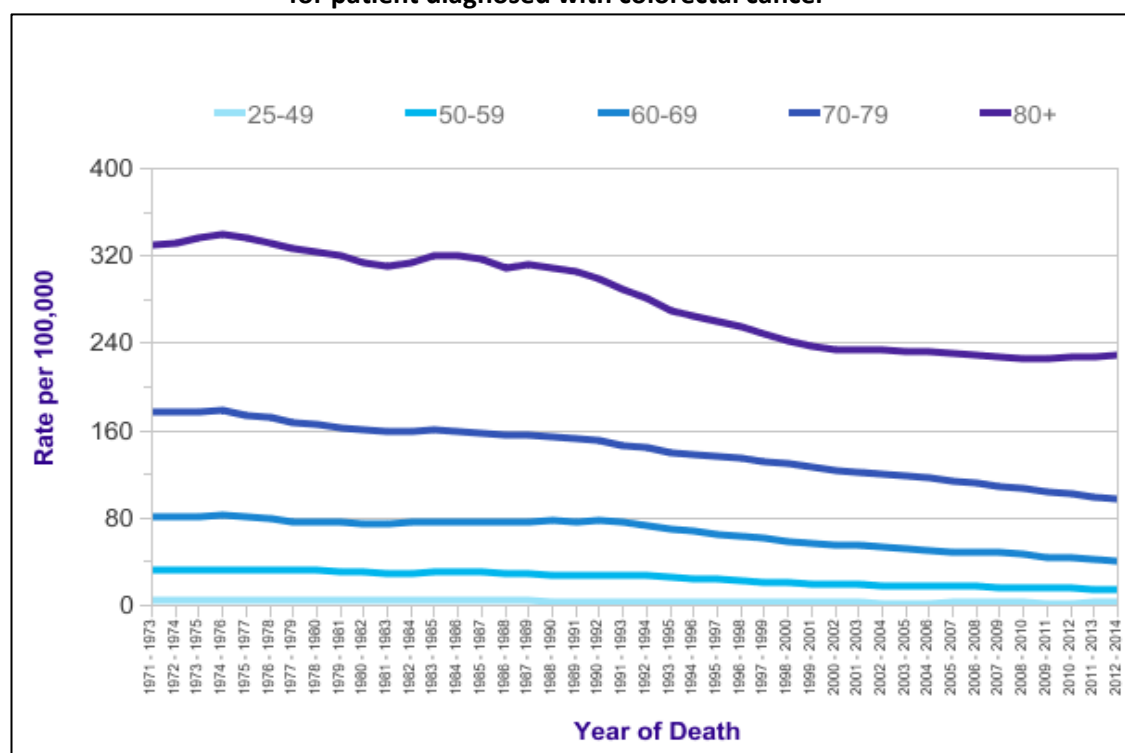
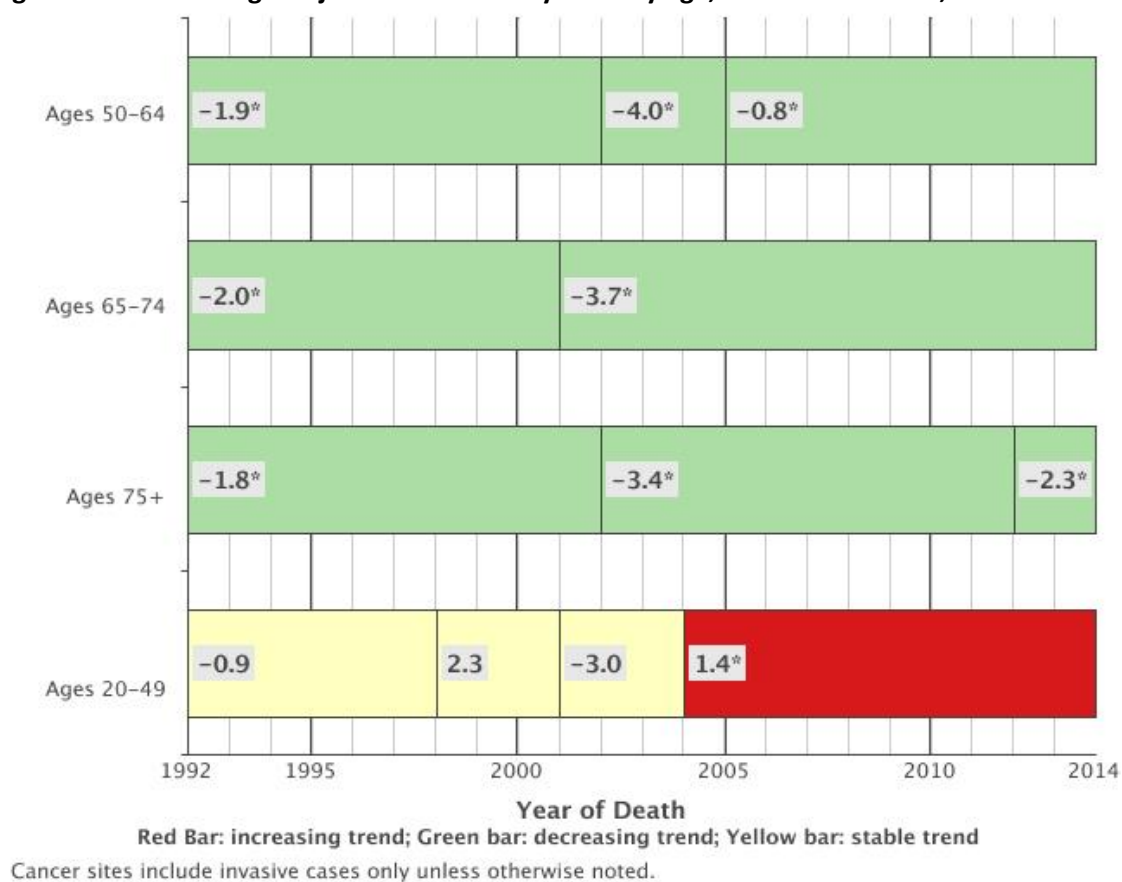


Figure 9 - Trends in Age-Adjusted US mortality rates by age, Colon and rectum, 1992 - 2014¹¹²



The difference in outcome between the two countries could be explained by variation in the diagnostic pathway or differences in the treatment following diagnosis. However comparisons between the two countries is beyond the scope of my thesis.

Stage data according to age has been analysed using US data. There is no comparable analysis using UK data. Wang et al.¹¹³ used SEER data, to stratify 279 623 patients according to 20 – 40, 41 - 50 and over 50 years of age. There was a higher proportion of stage III and IV tumours in the youngest group (Stage III 32.5% vs 29.6% vs 24.4% respectively; Stage IV 24.4% vs 22.9% vs 18.6%). There was also a higher proportion of undifferentiated tumours in the young groups. Overall the 3 year cumulative CRC specific survival was 71%, 74.0% and 69.4% respectively. The 5 year cumulative specific survival was 65.1%, 67.1% and 62.8% respectively. When stratified according to stage, patients over 50 had worse 3- and 5- year CRC specific survival than those aged under 40 ($p < 0.001$) and 40 – 50 ($p < 0.001$) at stage I. The same trend for both 3- and 5-year specific survival at stages II, III, and IV was noted (all $p < 0.001$). Compared with the youngest group, the patients ages 41- 50 had better 3- and 5-year specific survival at stage I ($P < 0.001$) and stage III ($P < 0.001$, 0.046), but not at stage II and stage IV. In Cox proportional hazards modelling, the older age group remained an independent predictor of poor prognosis.

Kim et al.¹¹⁴ analysed colorectal patients who had undergone a palliative or curative resection in a South Korea tertiary centre. 693 patients aged 45 and younger were compared with 1823 patients aged 56 to 65. Undifferentiated and advanced stage (III or IV) cancers (55.6% vs 47.9%) and rectal cancers (48.6% vs 40.8%) were significantly more common in young patients. The recurrence rates of stage I and II in the young group was higher than the middle age-group (8.8% vs 2.7% $p < 0.0001$). For all patients, the 5-year cancer-specific survival rate was 81.2% in the young versus 87.8% in the middle age group ($p < 0.001$). However the Kaplan Meier and 5 year survival rates were not significantly different between the age groups if stratified according to stage.

In another study based on the SEER database by Abdelsattar et al.¹¹⁵, patients were split into those under 50 and older. There was a lower proportion of localised disease, defined as confined within the bowel wall, in young patients (35% vs 42.6%). After adjusting for

gender, race, marital status, tumor location and year of diagnosis, younger patients were significantly more likely to present with regional disease (local lymph nodes involved) than localised disease and even more likely to present with distant disease (metastases) than localised disease. Younger patients with distant disease were more likely to receive cancer directed surgery on their primary tumour compared to the older patients, and young patients with rectal cancer more likely to receive radiotherapy for all stages of disease. The overall 5 year crude survival was 67.7% for the younger age group compared with 66.8% for older patients ($p = 0.008$). Comparing for stage, and adjusting for sex, race, marital status, tumour location, surgical resection or radiation the results are notably different. The adjusted cancer specific survival was better in the younger patients, across all stage groups. The 5-year cancer specific survival for young patients was 95.1% versus 91.9% for patients aged over 50 ($p < 0.001$) for localised disease, 76% versus 70.3% ($p < 0.001$) for regional disease, and 21.3% versus 14.1% ($p < 0.001$) for distant disease, respectively.

Li et al.¹¹⁶, using data from China, also found a later stage at diagnosis for young patients, but similar survival after accounting for stage, surgical approach and grade.

2.6.3 Difference in presentation

An analysis of the routes of diagnosis data presented above demonstrated patients under 50 were most likely to present to the A&E (66% vs 61% for 60 – 69 year olds), and least likely to be referred by the GP (23% vs 28% for 60 – 69 year olds)⁹⁸.

In a systematic review by O’Connell et al.¹¹⁷, which included studies up to 2001, the most common symptoms found were per rectal bleeding, abdominal pain, weight loss and change in bowel habit in that order. There was no comparison with older patients.

Kim et al.¹¹⁴ demonstrated abdominal pain (18.5% vs 13.8%) and bleeding/anaemia (42.4% vs 32.1%) were significantly more common in the younger age group.

Chan et al.¹¹⁸ described the presenting symptom, stage and survival for young (under 50) and older patients who presented to Stanford university with a diagnosis of colorectal adenocarcinoma between 2008 and 2014 inclusive. Younger patients were more likely to

have rectal cancer. Colon cancer in young patients was more likely to present with abdominal pain (41.2% vs 27.2%) and rectal bleeding (28.8% vs 23.2%), but only abdominal pain was more common (15.0% vs 7.4%) for rectal cancer. There were too few patients with other symptoms to find differences. The median time to diagnosis (as defined from first symptom noted by the patient to diagnosis) was significantly longer for patients aged under 50 (median 128 days IQR 60 – 265) vs those aged 50 years and older (median 79 days IQR 31 - 184). Symptom duration (from first symptom noted by the patient to presentation) and work-up duration (from presentation to diagnosis) were also significantly longer for younger patients. However there was no significant association between longer symptom or work up duration and later stage, for either age group.

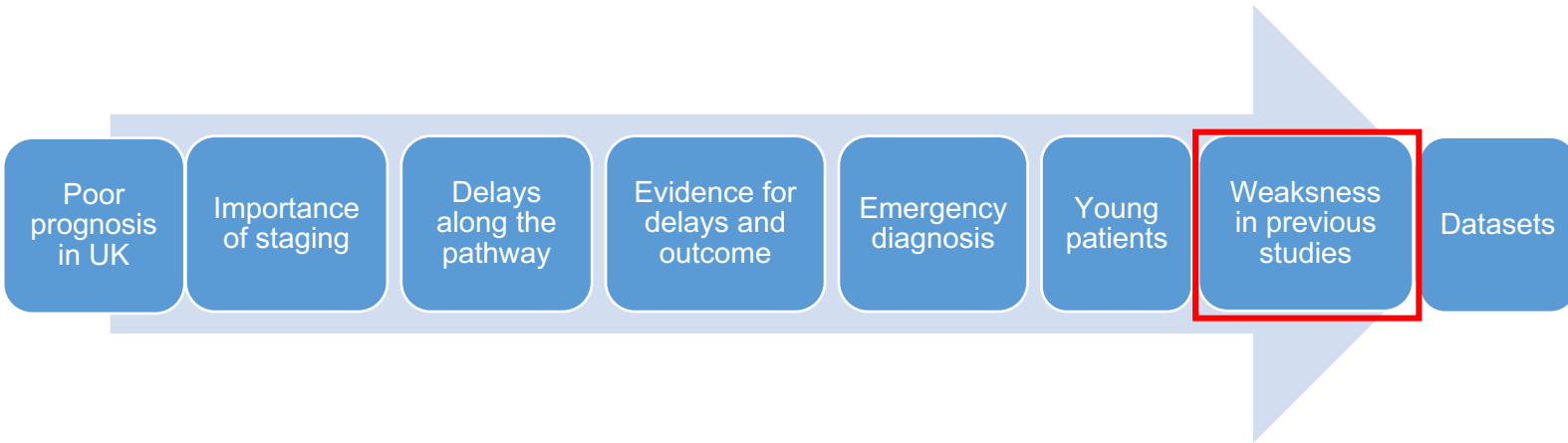
2.6.4 Summary of young patients and delays in diagnosis

The review of studies described above demonstrates young colorectal cancer patients are more likely to present with a later stage at diagnosis, and as an emergency. However after adjusting for confounders such as co-morbidities, stage and treatment, prognosis in young patients is similar to older patients. The main reason for this seems to be due to the more aggressive nature of treatment options made available to young patients. However it remains unclear whether young patients experience delays in primary care, and whether these lead to the findings above.

2.6.5 Why are young patients defined as under 50 in this thesis?

At present, studies concerning colorectal cancer in young adults demonstrate inconsistent findings including clinic-pathological features and survival outcome. This may partly be explained by the lack of an agreed definition of a young patients. Some research on “young” include the patients aged under 40 years old, whereas others have used age 50 as the cut-off. A structured review of studies up to 2003 revealed 37 out of 55 references defined “young” as patients <40 years old¹¹⁷. On the contrary, according to the Amsterdam Criteria, age 50 years is an important factor in the determination of patients with possible underlying hereditary predisposition to colorectal cancer such as Lynch syndrome¹¹⁹. Furthermore, some studies have recommend colorectal cancer screening to begin at age 50 years in average-risk individuals^{115,120}. To date, the number of studies using 50 years as cutoff age is progressively increasing. In practice, although age is often grouped, it represents a

spectrum, as demonstrated by the rates of colorectal cancer incidences describe above. As such categorization may lead to confounding results of young-onset and later-onset colorectal cancer. The cut-off age should also consider there are sufficient patients to power the relevant studies.



2.7 Weaknesses in previous studies to be addressed in this thesis

Clearly from the studies presented above, which span over 40 years of research, it is difficult to conclude with certainty that delays in diagnosis are associated with worse, or even improved, survival or stage. The discrepancies arise from difference in methodology in terms of patient inclusion and analysis, which are explained below.

1. Single center studies

- Patients are often identified from a single hospital or from one area in a country. These patients may not be representative of the general population in terms of demographics and co-morbidities. Also, the referral and diagnostic process may differ depending on the location of the hospital.

2. Insufficient patient numbers

- Due to point number 1, there may be insufficient patient numbers to find significant differences with longer delays.
- Subsequently studies tend to amalgamate rectal and colon cancer patients together. These tumour sites are known to present with different symptoms, with prevalence dependent on the patients age. The type of resection, and adjuvant treatment, also varies between the tumour sites.

3. Patients identified from hospital data only

- Patients identified from secondary care data may over represent those who are deemed medically fit enough to undergo a resection, or those that present as an emergency, leading to selection bias. By identifying patients from the cancer registry, the study population would include all cancer patients, irrespective of the mode of diagnosis or treatment.

4. Method of data collection

- Observational studies regarding delays by nature require retrospective data of the type and dates of symptoms before diagnosis. Questionnaires provided to primary care may be liable to selective reporting as GPs may under report the diagnostic interval if they felt the patient has come to harm. Alternatively, practices may fail to return questionnaires leading to a selection bias.
- Data that is not bespoke for the purposes of the study could lead to missing data. For example primary care notes may only include details of what the GP felt was important at the time of the consultation, which could differ to what the researcher would like to know.
- Patient surveys could lead to recall bias as they may falsely attribute a particular symptom to the cancer due to its proximity to diagnosis, thereby ignoring non-specific or less severe symptoms that would have occurred before.

5. Patient inclusion/exclusion criteria

- Patients diagnosed through an emergency are known to differ to non-emergency patients in terms of age, gender and co-morbidities. Emergency route should be included as a confounder or such patients should be excluded.
- Studies have excluded patients who die shortly after diagnosis. This has the potential to introduce selection bias by eliminating patients with a short diagnostic interval but higher mortality.
- Studies may fail to eliminate Death Certificate Only (DCO) patients, for whom the cancer was found incidentally at post-mortem. For such patients there were no opportunities to identify the cancer by the GP.
- Patients with a previous history of cancer, or conditions associated with cancer (for example inflammatory bowel disease), should be excluded from analysis as GPs are likely to be more aware of a cancer diagnosis in such patients, thereby reducing the interval to referral.

6. Definition of clinically important events

- The start of the diagnostic pathway has been defined as the symptom onset

or first presentation to primary or secondary care. When analyzing delays in primary care, the first presentation should be considered the start of the pathway.

- The definition of the end of the pathway also differs according to studies. Ideally this should be the date of diagnosis as confirmed by histology, which can be extracted from the cancer registry. Some studies have preferred to use the start of treatment as the end of the diagnostic pathway, which could erroneously elongate the diagnostic interval, but reduce the treatment interval.

7. Survival analysis

- Survival has been measured from the date of first symptom, date of diagnosis or date of treatment. When investigating the effect of delays in diagnosis, survival should be measured from the date of diagnosis. If survival is measured from the first symptom, the survival interval includes the interval of investigation. By using the treatment date, this would lead to lead time bias. A similar issue occurs if survival is measured from the last hospital attendance.

8. Confounding by indication

- When a patient presents to the GP, the doctor decides on the urgency of referral depending on the type of symptom and health of the patient. As such patients who are unwell at presentation, possibly due to a late stage tumour, are likely to be referred urgently. By analysing all colorectal patients together, studies often fail to recognize this issue, leading to confounding by indication. Studies have attempted to address this by analyzing patients with 'alert' symptoms separately to patients with non-specific symptoms. However the stage and patient interval is known to differ between symptoms grouped together as 'alert'. Alternatively each presenting symptom could be analysed separately, but this requires sufficient number to demonstrate a difference in outcome.

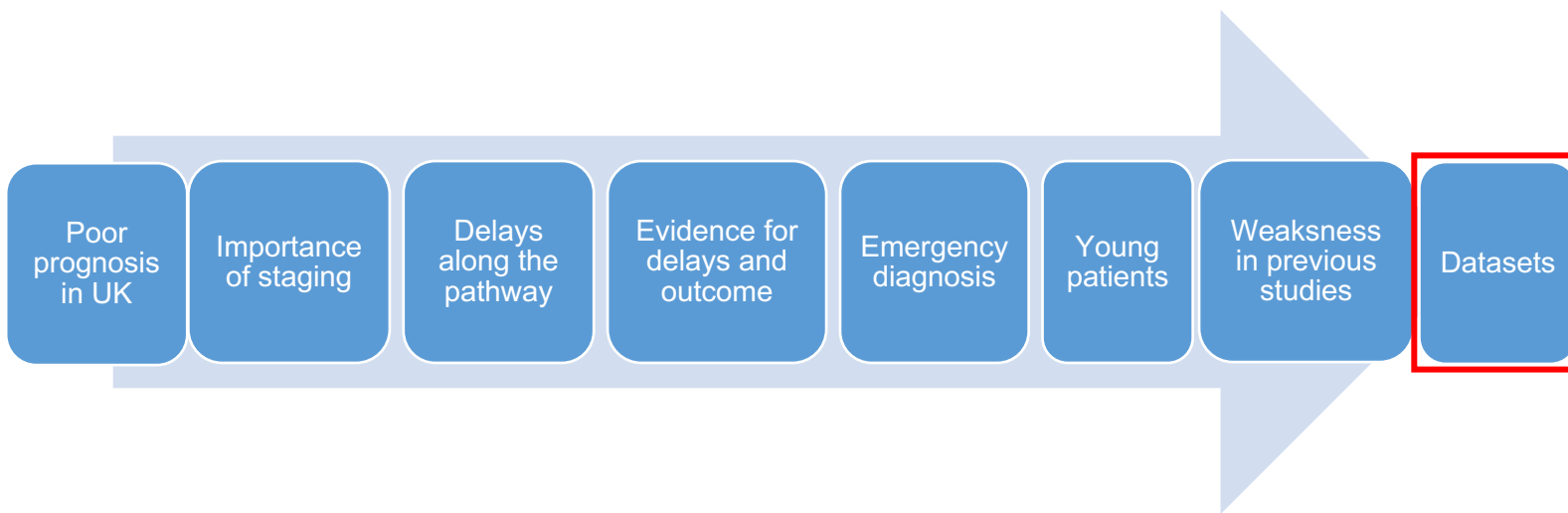
9. Patients factors as confounders

- Reasons for diagnostic delays and patient outcome are dependent on age, gender, co-morbidity and socio-economic deprivation. However due to the source of data, studies have failed to include these potential confounders.

10. Tumour characteristics as confounder

- Stage of tumour has been included as a confounder for studies determining survival following diagnostic delays. I believe stage should be considered a dependent variable in the investigation of delays, rather than a confounder.

To overcome these methodological issues, I have chosen to use the Clinical Practice Research Datalink as the source of patients for this thesis. The following chapter explains how the data in CPRD is collected, the linkage process to other national datasets, and its limitations and strengths.



2.8 Datasets used in this thesis – Clinical Practice Research Datalink, the Cancer Registry, Hospital Episodes Statistics and the Office for National Statistics

The hypothesis for this chapter:

CPRD linked data is a reliable source to investigate delays in primary care

The research questions are:

- Are cancer patients extracted from CPRD representative of the wider national cancer population?
- How does the diagnosis date in CPRD compare to the cancer registry?

The following chapter concentrates on the datasets I have used for my research. Initially I begin with a description of each of the four datasets, followed by previous publications that have investigated the accuracy of CPRD. Within this chapter I have included a narrative review that updates the findings from previous systematic reviews regarding the accuracy of CPRD. Within the results section, I describe three studies that determine the suitability of CPRD for my thesis.

2.8.1 The Clinical Practice Research Datalink

CPRD is a not-for-profit research service, jointly funded by the NHS National Institute for Health Research (NIHR) and the Medicines and Healthcare products Regulatory Agency (MHRA)¹²¹. CPRD provides anonymised primary care electronic records for public health

research¹²¹. From hereon, the primary care record dataset will be referred to as CPRD, although I acknowledge this is the name of the provider.

CPRD data is extracted from primary care providers using Vision software, covering approximately 8.5% of the UK population, equating to 5.69 million active patients and 13.3 million patients overall¹²². Patient interactions with primary care are recorded by the GP into the Vision database either based on a coding system or in the 'free text' section. The latter is no longer available to researchers. READ codes are used for consultations, with an alternative coding system for immunisations and prescriptions¹²¹. Vision requires a Read code to be selected for a clinical entry to be made into the patients medical records. Consultations are translated into a numerical medcodes, while prodcodes are used for prescriptions. Information recorded in Vision includes diagnoses, symptoms, prescriptions, referrals, tests, immunisations, health promotion activities, management plans and Quality Outcome Framework indicators.

Each practice is assigned a unique practice identifier, one of thirteen regions based on location, the last collection date indicating when the practice last downloaded patient data and the practice up-to-standard date which is when the practice was deemed to be of research quality (see below for criteria). For each patient, the registration date, transfer out date (if applicable), gender and birth year are provided. Within the consultation file, the reason for the consultation together with the 'system date' (date the entry was entered into Vision) and the 'event date' (date of the clinical event as recorded by the GP) are recorded. In the majority of records, these two dates are the same. In addition the consultations are categorized according to type, for example, surgery consultation, discharge details, telephone call, administration, night visit or prescriptions. Further clinical details are linked to 'look-up' files to add more information such as blood pressure recordings, weight and smoking. Referral data includes the reason for referral, date, urgency and the specialty receiving the patient. The 'Tests' file can provide the results of laboratory and bed side tests carried out by the GP, while the therapy file provides the name of medication and dosage.

Patients are deemed acceptable for the purposes of research¹²² if they have a

- valid gender

- birth year
- no events prior to birth year
- age of less than 115 at the last collection date
- first registration on or after the birth date
- current registration must be valid and a permanent registration.

If a patient has missing or invalid dates for all events, then the patient is also deemed unacceptable. On the practice level, the data is considered of research standard if practice mortalities are within an expected range and if there is continuity of recording within a practice. The former ensures that data is provided for patients who died rather than simply erased from the system, while also being a marker of irregularities in practice.

Within this thesis, only patients deemed 'acceptable' from practices with UTS dates before the diagnosis date were included. By only including acceptable patients from UTS practices, the number of patients reduces to 4.4 million active in the original CPRD dataset, representing 6.9% of the UK population, with a further 6.9 million records for inactive patients¹²².

It is possible to link CPRD to inpatient Hospital Episodes Statistics (HES), Index of Multiple Deprivation (IMD), Office for National Statistics (ONS) mortality data and the National Cancer Registration and Analysis Service (CR)¹²³. Data is only available for English practices providing consent for linkage, which at the time I received the data was 375, representing 55% of practices, thereby further reducing the number of patients available for research to around 4% of the population. Linkage is based on patient NHS number, date of birth, gender and post code. CPRD and the custodian of the linked data separately provide these details to Health and Social Care Information Centre (HSCIC) who carry out the process of linkage. The data is returned to CPRD fully anonymized. The coverage for each dataset varies: CPRD – 01/04/1990 to 01/04/2016 HES – 01/04/1997 to 30/09/15, ONS mortality – 02/01/1998 – 30/09/15, CR – 01/01/1990 to 31/12/2013. The individual datasets provide similar information as if the datasets were provided alone. Further description regarding the data fields, and the process of cleaning the data, is provided below.

My study protocol was reviewed by the Independent Scientific Advisory Committee (ISAC) as per the guidelines before data acquisition. As there was cancer registry linkage, NCIN also reviewed the protocol, which is available in the appendix (Appendix i).

2.8.1.1 Strengths and weaknesses of CPRD

CPRD provides a potentially large cohort of patients with a relatively long median follow up of 9.4 years (IQR of 3.4 – 13.9) for active patient and 5.1 years overall (IQR 1.8 – 11.1 years)¹²². As potentially every consultation is recorded through Vision, CPRD provides a wide range of possible conditions for analysis, including those with a low prevalence, which otherwise would have been difficult to analyse through bespoke collected data. CPRD is representative of the UK population in terms of gender, age, ethnicity and Body Mass Index, although it may under represent certain geographical areas, such as the north east of England and mortality rates, especially in young patients¹²¹. Data quality is particularly strong for conditions under the Quality and Outcomes Framework. CPRD is routinely collected, so the cost of acquiring this data in terms of money and time is significantly less than if data was collated bespoke for a study, especially if high number of patients were required. However, the routine nature of data collection means data will be missing¹²¹. As mentioned previously, a consultation can only be stored if a Read term is provided by the GP. However for a presenting symptom, the detail regarding when the patient first felt the symptom, and whether there were a combination of symptoms at presentation, may be recorded in the free text which is not available to researchers. The absence of a Read code for a disease or symptom is interpreted as absence of the symptom or disease itself. This produces a high positive predictive value, but reduces the sensitivity¹²⁴. The amount of free text is dependent on the recording style of the GP, making it difficult to use imputation methods to replace missing data in terms of symptoms. For each study the authors are required to extract relevant consultation based on the Read codes of interest. This can vary between studies, resulting in different results despite using the same source.

2.8.1.2 Systematic reviews into the accuracy of diagnosis recording in CPRD

Herret et al.¹²⁵ included 212 publications in their systematic review which found the median proportion of cases with a confirmed diagnosis overall was 89% (range 24% to 100%). Eighty-three percent of validations were external, i.e. with the use of a GP questionnaire,

request of medical notes or comparison of prevalence rates to other English based datasets. The remaining validations were internal (review of free text or diagnostic algorithm using medications and/or presence of symptoms to validate a diagnosis). The percentage of cancer cases confirmed rose to a median of 95.5% (Range 74 – 100). However the relatively poor quality of the studies meant the authors were unable to comment on the sensitivity and specificity of diagnosis. In an alternative systematic review, Khan et al.¹²⁶ restricted the study inclusion to comparisons of diagnosis recording in CPRD to the gold standard as defined by the authors– GP questionnaire, primary care medical records or hospital letters. They found the positive predictive value was particularly low (>50%) for acute conditions, such as liver injury, pancreatitis or renal failure. A PPV of over 90% was achieved for anorexia, congenital heart defects, inflammatory bowel disease and diabetes amongst others. Of note, studies with cancer diagnoses were not included, nor the accuracy of symptom recording. Only inflammatory bowel disease, myocardial infarction and dementia were analysed in terms of differences in the date of diagnosis.

Using similar definitions as Herret et al.¹²⁵ I carried out an updated systematic review (together with a fellow researcher at the time, Osama Moussa), to determine the accuracy of CPRD. The search strategy is described in the Appendix (section ii). A total of 1720 abstracts were sourced from the Medline, EMBASE and website searches, of which 927 were not CPRD studies following review of the title and abstract. Furthermore, reviewing articles and thorough search of related journals and conference proceedings produced a further 310 studies. Overall a median of 86% diagnoses, with a range 24–100%, were confirmed with either internal or external information. However, the proportion of cases varied according to disease being investigated, although the median proportion was greater than 86% for majority of the conditions. This search revealed only two studies that compared the accuracy of cancer diagnosis in CPRD with the cancer registry, both of which are described below. In more recent years, the cancer registry can be considered the gold-standard for identifying cancer patients due to the use of multiple sources to reduce the number of missing cases.

2.8.1.3 Accuracy of cancer diagnosis date

Two previous studies have attempted to determine the accuracy of cancer diagnosis in

CPRD and the difference in diagnosis date. Dregan et al.¹²⁷ compared the recording of four cancers (lung, colorectal, gastro-oesophageal and urinary tract) between CPRD (or GPRD as it was known at the time of the study) and cancer registry as the 'gold-standard'. All patients had to have an 'alarm symptom' recorded in CPRD for inclusion. There were 5429 cancer diagnoses in GPRD and 5710 in the CR, with 5216 (91% of CR total) diagnosed in both sources. There were 494 (9%) diagnosed in CR but not in CPRD and 213 (4%) that were diagnosed in GPRD but not CR. The PPV of a CPRD cancer diagnosis was 96% for lung cancer, 92% for urinary tract cancer, 97% for gastro-oesophageal cancer and 98% for colorectal cancer. The sensitivity of CPRD, using CR for reference, for the four cancer diagnoses ranged from 85% for urinary tract cancers to 94% for lung cancers. CPRD diagnoses of all four cancers were associated with a high specificity (99%). The diagnosis in the CR was recorded for a median 11 days earlier than CPRD. The median difference in dates ranged from 4 days (IQR -22 to 13) for urological cancer to 3 days for lung (IQR -31 to 6).

Boggon et al.¹²⁸ analysed the cancer diagnosis between CPRD and the CR for twenty-two cancers in patients who were initially extracted for a diabetes related study. They analysed the free-text as well as the Read terms to identify a cancer diagnosis, and assumed the presence of cancer therapy indicated a cancer diagnosis. Of the 5797 cancer cases in the CR (of which 681 colorectal, 465 oesophagogastric), 4830 (83%) were also recorded in CPRD. The highest non-agreement in cancer diagnosis was seen for older patients (OR 2.23 (95% CI 1.61–3.35) for patients aged 75+ years compared with patients aged <65 year), although no difference in gender was noted. The majority of patients diagnosis date in CPRD and CR were within one month of each other (63.9% of CPRD cases confirmed in the CR). For each cancer, survival at both 1 year and 5 years was higher in those cases known to CPRD than in those known to the CR. The differences in 5-year survival were considerably smaller than for 1-year survival, ranging from 0.4% (breast cancer) to 2.8% (melanoma).

2.8.1.4 Summary of the accuracy and suitability of CPRD

These previous studies and our updated systematic review demonstrate CPRD is relatively accurate in the recording of cancer and non-cancer diagnosis compared to the CR. However I feel it is necessary to compare the diagnosis date and difference in demographics between the datasets for the following reasons:

- the studies do not sufficiently clarify the difference in diagnosis dates between the CR and CPRD for the purpose of studying delays
 - GP is likely to learn of a cancer diagnosis after they receive information from the hospital, so the diagnosis date in CPRD would be expected to be recorded later than the cancer registry
 - A reliance on CPRD to determine the diagnosis date, as has been carried out in previous studies regarding delays, may erroneously describe a prolonged diagnostic interval
- there is no comparison of diagnosis dates between HES and CPRD or the CR
 - A comparison of all three will allow which dataset is likely to provide the earliest date of diagnosis
- by including patients only with alarm symptoms, or using a dataset extracted for a diabetes study, the findings described in the Boggon et al.¹²⁸ and Dregan et al.¹²⁷ study may not be relevant to the general cancer population. 40% of cancer patients may present with non-specific symptoms
- there is no mention if patients with a later diagnosis in CPRD differ to the remaining patients
- CPRD may exclude patients diagnosed as an emergency if a patient never presented to primary care before diagnosis, and died shortly after

2.8.1.5 Previous use of CPRD for colorectal cancer symptoms

Previous CPRD based studies provide a guidance of the expected proportion of patients with each symptom (table 7). The Hamilton et al.⁴³ and Sheringham et al.⁹⁷ studies were based on primary care records, included patients with a confirmed diagnosis in the cancer registry and described symptoms within a year leading up to cancer diagnosis, not just at presentation. However despite these similarities, there is a discrepancy in the percentage of cancer patients with a particular symptom. For example 42.4% of colorectal cancer patients in the Hamilton et al.⁴³ study experienced rectal bleeding as opposed to 17.9% in the Sheringham et al.⁹⁷ study. The difference in year or area cannot reasonably explain the wide difference. Instead the medical coding system is likely to be responsible - Hamilton et al.⁴³ used the International Classification of Primary Care, while Sheringham et al.⁹⁷ utilised the READ code system. The latter system is the basis of CPRD data, which would explain the

relatively similar findings from Renzi et al.¹⁰⁵. These studies provide a comparison, to ensure I have captured a similar proportion of patients and symptoms.

Table 7 Prevalence of symptoms in the year leading to colorectal cancer diagnosis

	Pruitt ⁷⁵		Renzi ¹⁰⁵				Redaniel ⁸¹	Hamilton ⁴³	Stapley ⁷⁷	Sheringham ⁹⁷
Number	6702	2976	1029	577	668	490	5912	349	349	943
Data source	SEER		CPRD				CPRD	PCR	PCR	PCR
Route	Excludes emergency		All		Excludes emergency		All	All	All	All
At presentation	Yes		No		No		Yes	No	Yes	No
Site	C	R	C	R	C	R	CR	CR	CR	CR
Rectal Bleeding	20.7	41.6	9.4	21.5	12.6	23.7	20.7	42.4	28	17.9
Abdominal pain	19.3	13.0	25.1	9.4	25.5	8.8	28.0	42.4	24	16.2
Constipation	4.6	4.2	8.1	8.2	8.7	8.6	11.5	26.1	23	17.0
Anaemia	33.7	18.2	19.2	6.2	23.2	5.9	-	27.3	24	50.2
Diarrhoea			9.9	12.0	9.9	11.2	14.9	37.8	22	9.0
Change in bowel habit			5.5	11.6	6.7	12.2	13.7	-	-	
Abnormal stool	4.7	3.8						-	-	
Weight loss only	4.6	5.1	3.1	1.7	3.1	1.6	3.5	26.9	7	
Fatigue	15.7	11.7	4.4	2.3	4.9	2.5	4.6		-	
Abdominal mass							2.5		-	
Weight loss or fatigue										2.8
Other (rectal mass, flatulence, altered bowel habit)									-	10.8

PCR - Primary care records, C = Colon, R = Rectum, CR = Colorectal

2.8.1.6 Validation of CPRD as a data source to study delays in clinical pathway

The accuracy and completeness of CPRD is key to investigate delays in primary care, and whether the findings of this thesis can be applied nationally. Below I briefly explain the aims and the studies used to investigate these.

1) Completeness of cancer diagnosis in CPRD

a) Does CPRD capture all the cancer cases?

- If patients are being missed then CPRD does not reflect the general cancer population and therefore reduces its ability to influence health policy

2) Accuracy of cancer diagnosis date

a) Does the cancer diagnosis date in CPRD correspond to the diagnosis date in the cancer registry?

- Investigation of delays along the patient pathway requires an accurate diagnosis date

These two points are discussed through the presentation of two studies I carried out comparing the recording of cancer diagnosis and date between CPRD, HES and the CR and comparison of patient demographics between patients extracted from CPRD and national registry data.

3) Can delays in CPRD be reliably related to other studies?

- a) Delays in adjuvant chemotherapy has been analysed previously using HES – does CPRD, together with the linked HES data, provide similar results? This will allow accuracy of lost to follow up, complications and ONS mortality data.

2.8.2 Hospital Episode Statistics

The linked HES dataset provided by CPRD is similar to the larger original HES dataset that can be requested alone, except it only includes routinely collected inpatient episodes for the study patients. Each episode contains a diagnosis code based on ICD 10 coding and a procedure based on OPCS 4 coding if applicable. Over the years, HES has become a robust, high quality source of secondary care data, that has been verified for various conditions and procedures¹²⁹. Relevant to my thesis, with HES it is possible to determine the route of cancer diagnosis, adjuvant treatment, resections, complications, and re-admissions.

2.8.3 Cancer registry

The linked cancer registry is provided by Public Health England (PHE) via the National Cancer Registration and Analysis Service (NCRAS) who co-ordinates the national collation from 500 different local and national feeds within England to accurately describe the diagnostic and treatment pathways for cancer patients¹³⁰. Once all the data has been collected, 40 various vital checks are carried out before publication, such as the incidence date is not after the date of death^{9,10}. In recent years, only 0.1% of registrations failed at least one vital test¹³⁰.

2.8.3.1 Accuracy

Ideally the majority of diagnoses should be based on histological sampling, with a hierarchical source of diagnosis, as described below (table 8)¹³⁰. Code 'seven' indicates the preferable source.

Table 8 - Source of cancer diagnosis as defined by the Cancer Registry¹³⁰

Code	Description	Criteria
0	Death certificate only	The only information to the registry is from a death certificate.
Non - microscopic		
1	Clinical	Diagnosis made before death, but without the benefit of any of the following (2–7)
2	Clinical investigation	To include all diagnostic techniques, including X-ray, endoscopy, imaging, ultrasound, exploratory surgery (e.g., laparotomy) and autopsy, without a tissue diagnosis.
4	Specific tumour markers	To include biochemical and/or immunological markers which are specific for a tumour site
Microscopic		
5	Cytology	Examination of cells whether from a primary or secondary site, including fluids aspirated using endoscopes or needles. Also to include the microscopic examination of peripheral blood films and trephine bone marrow aspirates.
6	Histology of a metastasis	Histological examination of tissue from a metastasis, including autopsy specimens.
7	Histology of a primary tumour	Histological examination of tissue from the primary tumour, however obtained, including all cutting techniques and bone marrow biopsies. Also to include autopsy specimens of a primary tumour.
9	Unknown	

2.8.3.2 Completeness

The cancer registry accuracy has been checked against patients records to reduce the number of missing patients, with the most recent figures from ONS suggesting completeness approaches 99%. This level of accuracy corresponds to registrations, rather than stage data, which is lower. The number of cancer registrations based on death certificates (DCO) is another measure of accuracy⁹¹⁰. If there is no previous mention of cancer within the available patient history, then the diagnosis date is taken as the date of death. These 'zero survival' situations may be correct for some patients. However a high rate of DCO within a cancer registry indicates a high reliance on death certificates, and suggests other sources of data are not being appropriately utilised. This will lead to missed cases. Relying on DCO may also place the diagnosis date into the wrong year, thereby producing inaccurate incidence rates.

2.8.3.3 Diagnosis date

There is a hierarchical priority of determining the incidence date of cancer, as suggested by the International Agency for Research on Cancer (IARC)⁹, which is followed by the NCRAS:

1. Date of first histological or cytological confirmation of this malignancy (with the exception of histology or cytology at autopsy).

This date should be, in the following order:

- a. date when the specimen was taken, or
 - b. date of receipt by the pathologist, or
 - c. date of the pathology report
2. Date of admission to hospital because of this malignancy.
 3. When evaluated at an outpatient clinic only: date of first consultation at the out-patient clinic because of this malignancy.
 4. Date of diagnosis, other than 1, 2 or 3.
 5. Date of death, if no information is available other than the fact that the patient has died because of malignancy.
 6. Date of death, if the malignancy is discovered at autopsy.

Although it is expected the majority of diagnosis dates will be based on number 1, it is not possible to ascertain the proportion from the dataset provided. The cancer registry is a dynamic dataset, so cancer registration details may be updated up to a year after the initial registration. In some cases a registration may be deleted, although this is uncommon¹³¹.

2.8.3.4 Variables available from the cancer registry

The following data fields were provided with the linked cancer registry:

- CPRD Patient identifier – allows linkage between the datasets
- CR patient identifier – a small proportion of patients may have more than one cancer registration for the same cancer, in particular if a patient moves from one practice to another. For my thesis, I analysed the first registration for the cancer of interest.
- Age – it is recommended the age should be obtained from the CPRD dataset

rather than the cancer registry

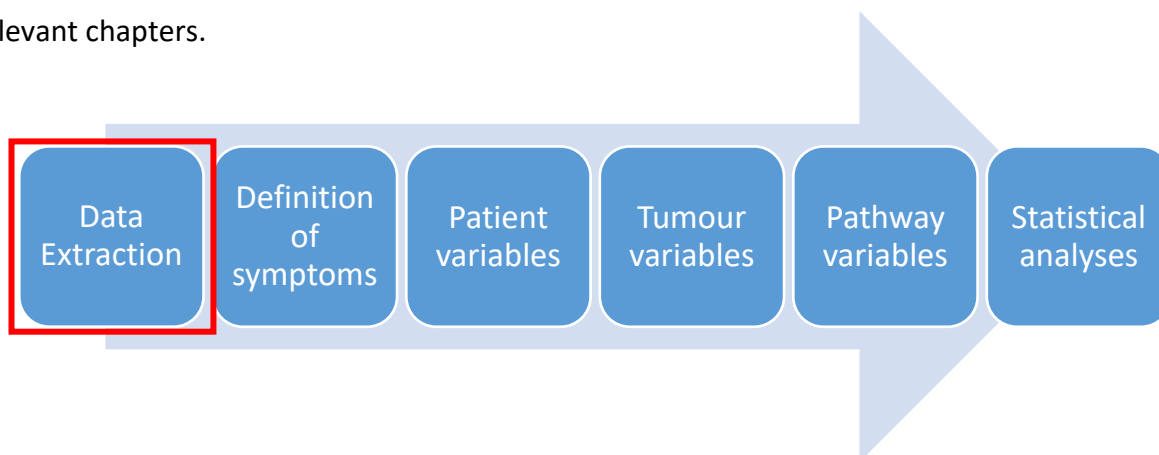
- Diagnosis date – based on the criteria above, provided in a day, month and year format.
- Basis of diagnosis – based on the criteria above
- Site of cancer – based on ICD 10 coding
- Laterality for paired organs
- Nodes excised or involved
- Screen detected
- Death Certificate Only – separate from the basis of diagnosis variable
- Hormone receptor positivity depending on cancer
- Duke Stage for Colorectal cancers if available
- Separate T, N, M staging
- Stage (Best) - Combined TNM stage
- Separate TNM for clinical staging
- Separate TNM for staging based on imaging
- Separate TNM for staging based on pathology

2.8.4 Office for National Statistics Mortality data

The cause of death, if applicable, is provided through linkage provided by ONS. This holds the underlying cause of death, as well as contributory cause of death, labelled 1 to 15¹³². The underlying cause of death is determined through an algorithm to better understand the condition that led to death, rather than simply based on level I on the death certificate. For example, a patient with a venothromboembolic event noted in line I of the death certificate may have an underlying colorectal cancer. In the ONS dataset the underlying cause of death is given as colorectal cancer, rather than the VTE¹³². Therefore it is possible to differentiate between cancer and non-cancer related cause of death. ICD-10 coding for deaths post 2001 and ICD 9 for earlier registrations, was used. Date of death was provided in day, month, year format.

CHAPTER THREE - METHODS

Within the following chapters there are variables which are common to each chapter. These are explained below. Methods particular to a study are explained in more detail in the relevant chapters.



3.1 Extraction of cancer patients from CPRD

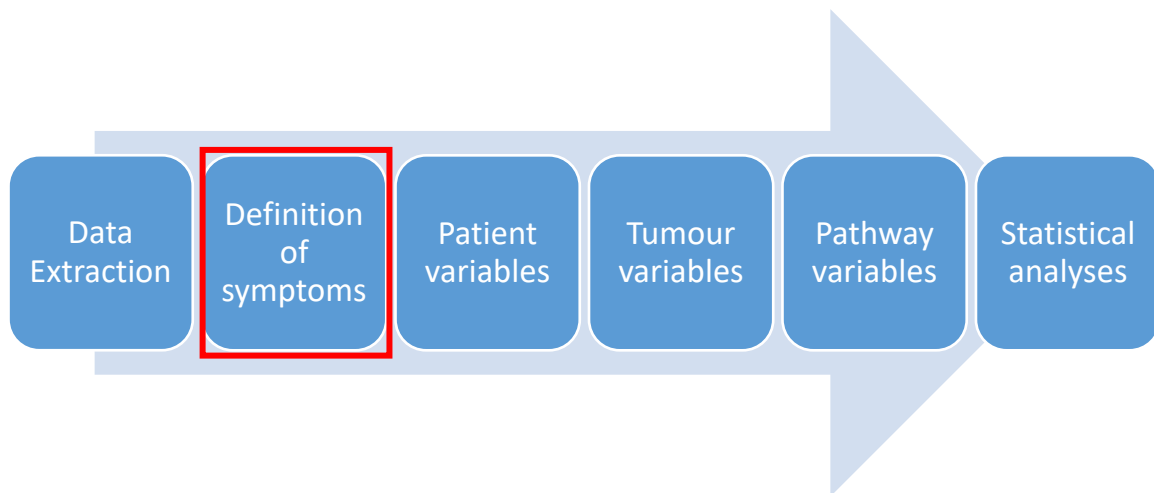
The protocol for my data request is set out in the Appendix (section i). In summary, the study population was identified from 'Set 12' which includes patients from GP practices who have given consent for linkage to CR, HES and ONS. From the CR, I requested all patients with a ICD 10 code of the cancers of interest – colorectal, oesophagogastric (OG), lung, prostate and breast (see appendix iii for relevant codes). For the remaining patients (i.e. no relevant cancer diagnosis in the CR), the HES dataset was searched for patients with a cancer diagnosis based on ICD 10 coding. The final search extracted patients from CPRD with a relevant cancer diagnosis, who by default did not have a cancer diagnosis in the CR or HES. Through this process the dataset contains patients with at least one relevant cancer diagnosis in CR, CPRD or HES. All HES and CPRD episodes for such patients were provided. The number extracted is as follows:

A total of 347,515 patients had a record in at least one data source.

Of these:

- 347,515 patients were 'acceptable'
- 346,970 patients were over the age of 18 at the date of diagnosis
- 194,047 patients were registered in an up-to-standard (UTS) practice (followed-up) on the date of at least one cancer diagnosis
- 177,376 patients have at least one year (365.25 days) of follow-up prior to at least one cancer diagnosis
- 177,376 patients were included in the extraction cohort
- 151,371 patients after exclusion of those with a missing diagnosis date, or duplicate cancer diagnoses. Patients with a mesothelioma diagnosis in CPRD were also excluded as this condition was not considered a lung cancer in analysis

30773 colorectal, 12944 OG, 41515 breast, 31951 lung and 34189 prostate cancer patients were identified with a diagnosis in either HES, CPRD or the CR with complete dates. The inclusion criteria for each study is described in the relevant chapters.



3.2 Definition of symptoms

Only patients with confirmed colorectal cancer were included in my thesis. Table 8 describes the symptoms that I considered relevant to colorectal cancer, which were based on the NICE guidance of 2005. These symptoms, and the grouping described, is similar to a study by Neal et al.⁵², which used CPRD to demonstrate a reduction in the diagnostic interval following the introduction of the 2005 NICE guidance. Renzi et al.¹⁰⁵ also used these symptoms to describe the likelihood of an emergency diagnosis. The Read codes for each symptom are provided in the appendix (see Appendix iv).

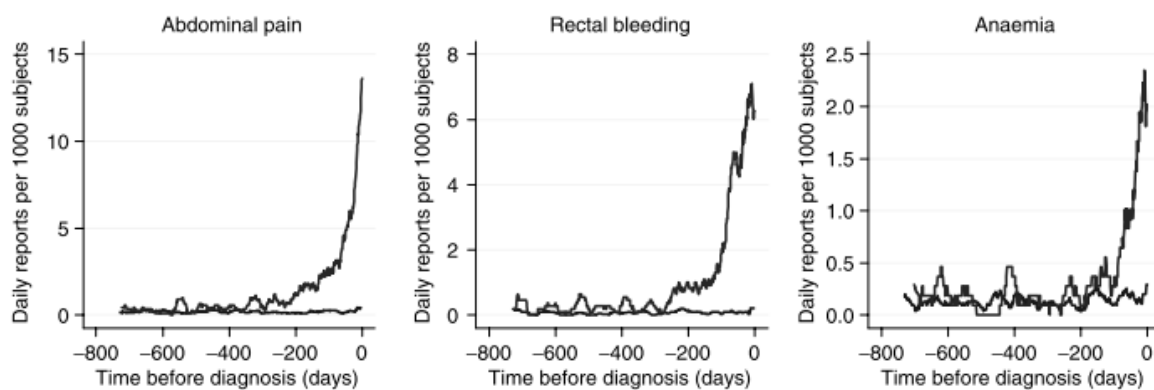
Table 9 - Symptoms considered relevant for colorectal cancer for this thesis

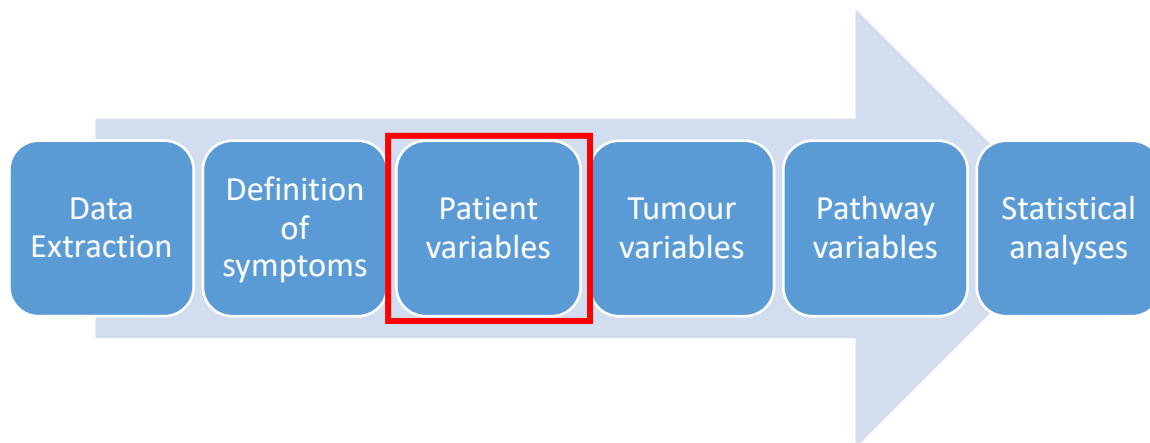
Red flag	Non-specific
Rectal bleeding	Abdominal pain
Change in bowel habit	Constipation
Diarrhoea	Weight loss
Abdominal or rectal mass	Fatigue
Anaemia	Other bowel function

The optimum interval before diagnosis to search for cancer associated symptoms is debateable. If the period is too short, episodes related to cancer may be overlooked. If the period is too long, symptoms related to benign conditions may be incorrectly linked to the eventual cancer diagnosis. It is impossible to determine this period with precision, especially in observational studies. Alternatively, it is possible to compare primary care usage for each symptom between benign and cancer patients. Hamilton et al. (2005)⁴³ found the consultation rate (daily reports per 1000 subjects) for colorectal cancer patients with abdominal pain or rectal bleeding would increase at about a year leading up to diagnosis, a trend not seen for benign patients (figure 10). For anaemia the rate would increase around

200 days before diagnosis. From this study it seems reasonable to use one year before cancer diagnosis as the threshold. A similar criteria has been used in previous studies using CPRD to investigate delays in diagnosis^{43,81}. In the chapters which include the point of referral as the key determinant along the patient pathway, I have used the year leading up to the referral date as the primary care physician would be reviewing the consultations leading up to the decision to refer.

Figure 10 - Primary care usage in the two years leading up to cancer diagnosis for benign (lower line) and cancer patients (upper line) ⁴³





3.3 Patient variables

Based on my findings described in the previous chapter, I believe it is important to adjust for the following factors in the investigation of delays:

- Symptom
- Age
- Gender
- Co-morbidities
- Socio-economic deprivation
- GP practice
- Number of consultations leading up to diagnosis
- Stage at diagnosis (depending on aim)
- Route of diagnosis

3.3.1 Age

Age is calculated from the date of birth provided in the CPRD dataset. Age will be analysed as categorical data in the following groups: Under the age of 50, 50 – 59, 60 – 69, 70 – 79 and 80 and over. The threshold of 50 for the youngest group corresponds to previous studies, so allowing a direct comparison (explanation provided above). The upper threshold of over the age of 80 have more co-morbidities, and so are more likely to be managed palliatively if cancer is confirmed - they are less likely to undergo surgical resection with curative intent and less likely to be offered adjuvant treatment for similar stage disease as younger patients^{133,134}.

3.3.2 Co-morbidities

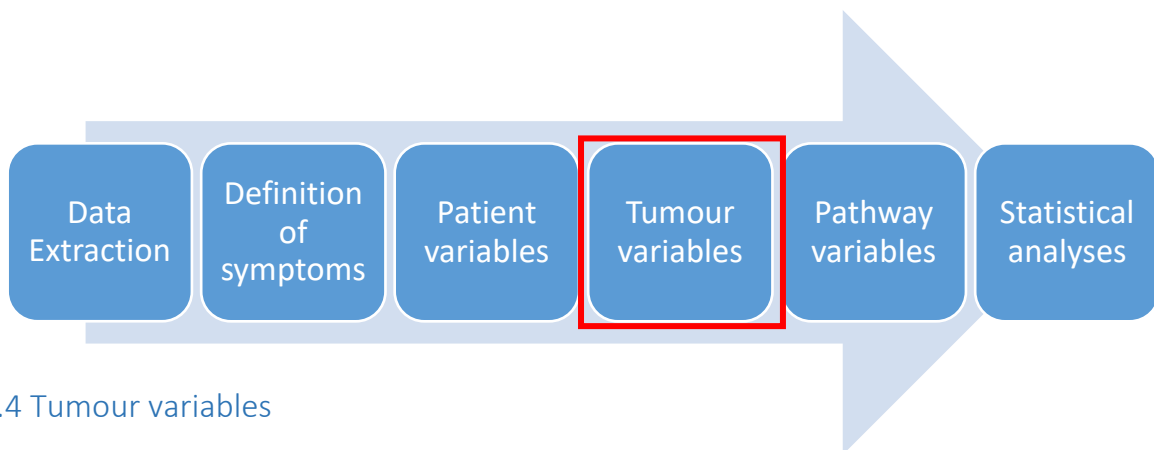
Co-morbidities will be measured using the Charlson Co-morbidity Index¹³⁵. As the CPRD data has been linked to HES, I am able to use both CPRD and HES as the source of diagnoses¹³⁵. Therefore conditions that are missed in the HES data should be captured by CPRD and vice-versa. The ICD codes and medcodes for co-morbidities can be obtained from the online version of a paper by Khan et al.¹³⁶. Three years of data leading up to the day before diagnosis will be used to identify co-morbidities. As only the first diagnosis of colorectal cancer in the cancer registry will be considered, and patients with any cancer previous to the cancer of interest will be excluded, Charlson scoring will not include cancer.

3.3.3 Socio-economic status

The Index of Multilevel Deprivation¹³⁷, as provided by CPRD, will be used as the metric for socio-economic status, based on the location of the patient.

3.3.4 GP practice

The GP practice, whether due to the number or demographics of the doctors working there, or a repeated pattern of patient management, has a bearing on the stage at diagnosis of cancer and the likelihood of an emergency diagnosis. For these reasons it is important to adjust for clustering within practices which if ignored will introduce bias. Therefore I will use the Generalised Estimating Equation function in SPSS.



3.4 Tumour variables

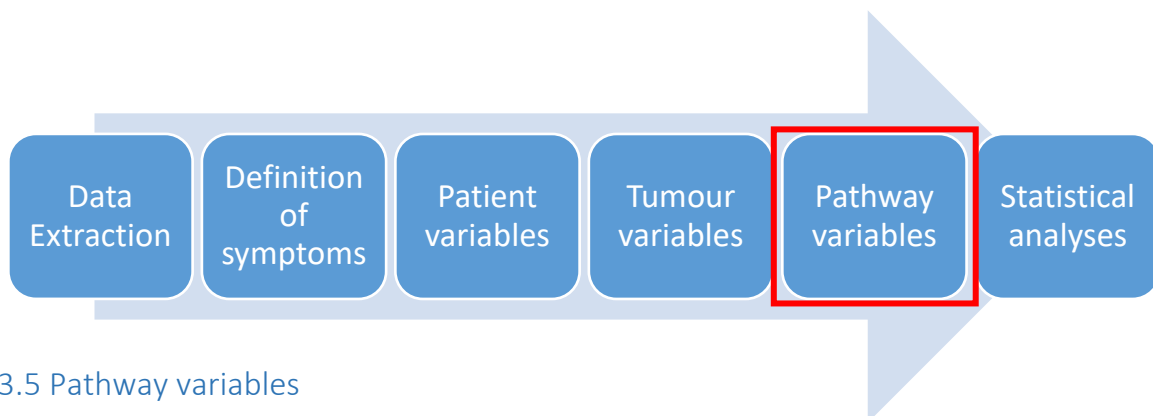
3.4.1 Stage at diagnosis

The 'best stage' variable provided in the cancer registry is a result of resolution between clinicians and pathologist at the multidisciplinary meeting (MDT). Although ideally this should be based on the pathology staging, it is not possible to know from the registry the process that lead to this staging. Researchers are recommended to use the separate TNM staging as there is no assurance that the "best stage has taken into consideration all available sources"¹²². Therefore I carried out an iterative process of cleaning the data to maximise the completeness of staging as described below and previously published by Walters at al.¹³⁸. A similar process is used by NCIN before the release of cancer statistics¹³⁹.

1. Stage was obtained from the combined TNM stage
2. If 1 not available, Dukes staging was converted to TNM staging (Duke A = stage 1, B = 2 etc.)
3. If Dukes and combined TNM stage not available, the tumour was deemed to be stage IV if evidence of metastases in pathology, clinical or imaging
4. For the remaining patients the stage was combined using the separate T, N and M staging. The highest of pathology, imaging and clinical staging was used
5. 'X' values and blanks were deemed missing data rather than considered as 'zero'. Any one of T, N or M missing meant a stage could not be provided, unless M positive (see point 3).

3.4.2 Site of cancer

Colon cancer site was based on the cancer registry ICD code of C18 – C19, while rectal cancer was C20.



3.5 Pathway variables

3.5.1 Route to diagnosis

The Route to diagnosis²⁹, which follows the algorithm as described above, was provided with the cancer registry data for patients with a registration between 2006 – 13 inclusive. As the cancer registry coverage starts from 1997, potentially nine years worth of cancers would be lost without an alternative method of determining an emergency diagnosis. Therefore I searched the six months of inpatient HES data preceding the cancer registry diagnosis date for the nearest episode with that cancer. If no HES data was found in this period, then the patient was deemed to have undergone a non-emergency diagnosis. This is a modification of the 'Routes of Diagnosis' study, as outpatient data was not available. In addition the work by Ellis-Brookes et al.²⁹ seemed to analyse any episodes, even if they had no cancer diagnosis associated with the episode. In my definition, the diagnosis of cancer had to be in the first three diagnosis variables. In most patients, colorectal cancer diagnosis is made after histological sampling of the cancer through an endoscopic examination or may be obtained after an operation. Both these clinical events are coded in inpatient HES, so I would expect the majority of patients to have an episode in HES related to the diagnostic process. Episodes after the cancer registration date would be related to complications of surgery or adjuvant treatment. A weakness with this method is cases diagnosed as an emergency following an outpatient HES end-point using the 'Routes to diagnosis' algorithm could potentially be labelled as a non-emergency as no outpatient was available. However, as shown above, only 6% of emergency patients are diagnosed in this way. This route is more common for Head & Neck cancers⁸⁷.

By restricting the colorectal cancer cases between 2006 – 2013, I found the overall sensitivity of this criteria to identify emergency cases as 72.1% and the specificity of 96.6%

(cases with an unknown, DCO or screening route were excluded) (table 10). Therefore for studies that concentrate on non-emergency routes, I could apply this criteria and use data between 2000 – 2005. However for studies involving emergency routes, I should restrict data usage to 2006 – 2013 inclusive.

Table 10 - Comparison of the Routes to Diagnosis and Emergency presentations within 6 months for colorectal cancer cases between 2006 - 2013 inclusive

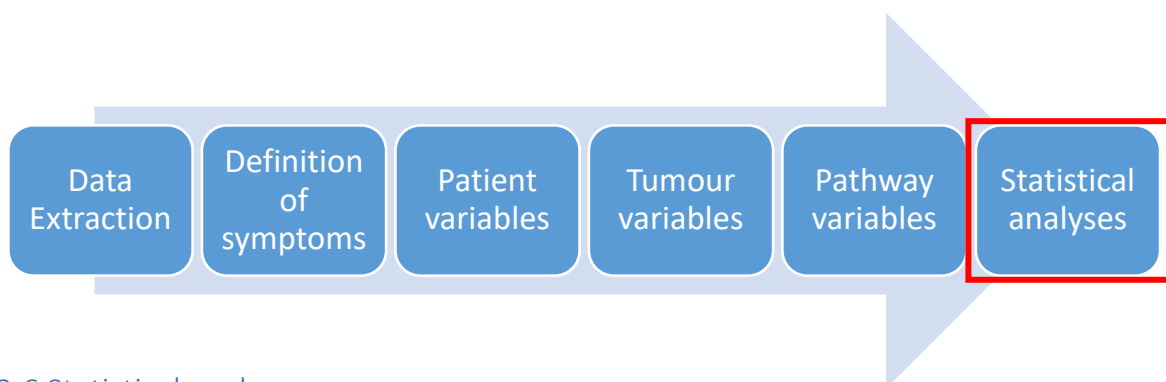
	Emergency within six months as determined by my algorithm	
	Yes	No
TWW	85	4005
GP referral	127	3123
Other outpatient	81	1030
Inpatient elective	15	592
Emergency	2377	922

3.5.2 Definition of referral

As the main interest of this thesis is delays in primary care, it is necessary to restrict the patients according to those with a referral from primary to secondary care due to the suspicion of cancer. The Read codes for this are presented in the appendix (section v). These were based on either the presence of a relevant symptom in the medcode associated with the referral, or a medcode with the destination or reason for referral, for example ‘Gastroenterological referral’ or ‘suspected lower gastrointestinal cancer’. Vague referral reasons, such as ‘referred to medical clinic’ were not included as the intention of the GP would have been unclear. There are no previous publications based on the referral code to allow comparison.

3.5.3 Type of consultations

Consultations will be labelled as ‘relevant’ meaning an associated symptom has been recorded whether it is a face-face consultation, home visit or over the phone consultations. Duplicate consultations, based on a consultation identifier occurring on the same day, will be counted once as it is unlikely a patient would see their GP more than once in a day.



3.6 Statistical analyses

3.6.1 Univariate analysis

All analysis was carried out in Statistical Package for Social Sciences (SPSS) v24 by IBM. Statistical significance was taken at $p < 0.05$, although where appropriate significance levels were characterized as necessary to lower levels to reduce the chance of a type I error. Chi-squared analysis was used for comparing proportions between categorical data. Linear by linear association compared categorical data where there was a natural order to the categorical data, such as the number of consultations. The Mann-Whitney U test compared the median values between non-normally distributed continuous data between two groups. If there were more than two groups, the Kruskal-Wallis test was used. The T-Test was used to compare mean for normally distributed data.

3.6.2 Multivariate analysis

Various regression modelling was employed depending on the study. All regression modelling was carried using the Generalised Estimating Equation (GEE) module in SPSS to take account of clustering in practices. For each study below I have stated the independent and dependent variables.

The following assumptions were checked.

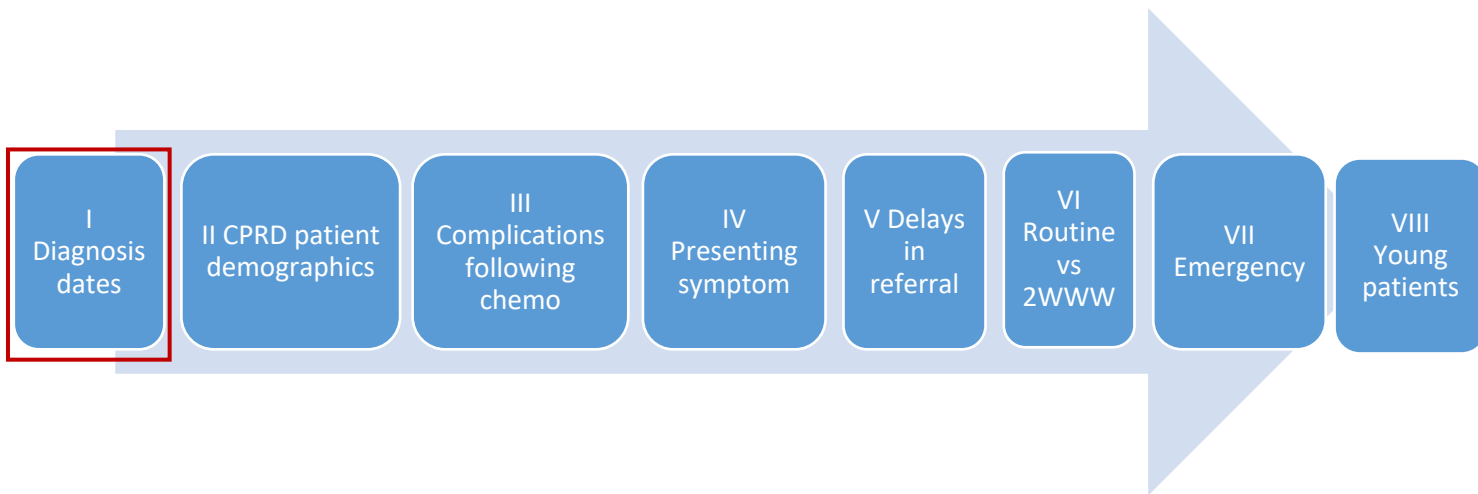
For binary logistic regression:

- The categories of the dichotomous dependent variable and the nominal independent variables are mutually exclusive, which is ensured by using the GEE module.
- There is no multicollinearity as checked against the Tolerance values
- There are no significant outliers

For survival analysis

- There is a similar amount of censoring in each group of interest
- Survival is measured from the date of diagnosis to reduce lead time bias
- For Cox modelling, there was no violation of the proportional hazards assumption
- Patients without a non-cancer related death in ONS were considered censored at the end of ONS coverage

CHAPTER FOUR - STUDIES



4.1 Study I - Comparison of cancer diagnosis dates between the Clinical Practice Research Datalink, Hospital Episodes Statistics and the Cancer Registry

In this chapter I compare the coverage of cancer diagnoses specifically and differences in diagnosis dates between the three national datasets

The hypotheses of this study states

- **Cancer diagnosis is recorded later in CPRD**
- **Relying on CPRD as the source for cancer patients leads to the exclusion of patients diagnosed as an emergency and/or with higher co-morbidities**

The aim of this study is to answer the following questions:

- **What is the coverage of cancer diagnosis between CPRD, HES and the Cancer Registry?**
- **Does the data source significantly alter the demographic profile of the study population?**
- **Which of the three datasets, CPRD, HES and the CR records the earlier diagnosis date?**
- **Which patients are more likely to have a later diagnosis recorded in CPRD?**

4.1.1 Introduction

Before robust statistical analysis can be carried out regarding delays, it is necessary to ensure the accuracy of key clinical events such as cancer diagnosis¹. Using the linkage described in this thesis, it is now possible to investigate discrepancies in coverage between datasets, and whether these discrepancies can lead to significant differences in patient characteristics. Time-to-event studies rely on the diagnosis date, so differences in this date by database are important; studies using CPRD often take the first recording of a given disease within CPRD as the date of diagnosis. These difference between dates for individual cancers has not previously been clarified^{127,128}.

The aim of this study is to describe the coverage of cancer diagnosis in CPRD, HES and CR, and whether utilising CPRD alone infers a selection bias. The magnitude of difference in diagnosis dates between the datasets, and whether patient characteristics or the route of diagnosis increases the risk of a later diagnosis in CPRD relative to CR, is also investigated.

4.1.2 Method

4.1.2.1 Patient inclusion

Patients included in this study had a diagnosis and date of diagnosis of colorectal, OG, breast, prostate or lung cancer diagnosis in one of CPRD, HES or the cancer registry. This was further refined to diagnoses between 2000 and 2013. For breast cancer only female patients were included. All patients were 18 years old or over, without an upper limit to the age.

4.1.2.2 Coverage and concordance of cancer diagnosis

For each cancer, I calculated the coverage provided by each dataset of all the patients identified, and the proportion of patients with a diagnosis in all three. In addition the concordance of cancer diagnosis was described, defined as the proportion of cases in a particular dataset covered by one of the other two sources. The age, gender, Charlson co-morbidity score, socioeconomic status, proportion of emergency diagnoses and cancer and non-cancer related mortality at one and five year was compared between patients identified from the CR, CPRD and HES using the Chi-squared test. These factors were defined

as previously described. Standardised residuals were compared to describe differences in the proportions of each variable between datasets at a significance level of $p < 0.05$.

4.1.2.3 Comparison of cancer diagnosis dates

The median difference between the diagnosis dates of two datasets was calculated in days, together with the interquartile range. Patients without a diagnosis date in both the datasets of interest and those with more than a year between diagnosis dates were excluded. The latter assumed only diagnoses recorded within a year represented the same cancer episode. The distribution of differences in cancer diagnosis dates between the datasets is demonstrated in the graphs below. In particular the percentage of patients with a diagnosis on the same day, and then earlier in either dataset is described.

Unadjusted analysis described the characteristics of patients who had a diagnosis in CPRD later than CR. A logistic regression model was used to determine the adjusted odds ratio of a later diagnosis in CPRD compared with CR for age, gender, Charlson score, emergency diagnosis and Index of Multilevel Deprivation (IMD).

4.1.3 Results

30773 colorectal, 12944 OG, 41515 breast, 31951 lung and 34189 prostate cancer patients were identified with a diagnosis in either HES, CPRD or the CR with complete dates. Of these, 24386 (79.2%), 10371 (80.1%), 30621 (73.8%), 25581 (80.1%) and 25811 (75.5%) patients respectively had a diagnosis between 2000 and 2013 in at least one of the datasets, and had at least one year of UTS data before the earliest diagnosis date. The latter cohort were included in the following analysis.

4.1.3.1 Coverage of all patients with at least one recorded diagnosis in either CR, CPRD or HES

For colorectal, OG, lung and prostate cancer, the CR contained the highest proportion of patients at around 90% (table 11). CPRD included more patients than CR only for breast cancer (94.4%), but provided the lowest coverage of CRC, OG and lung cancer. The proportion of patients in HES varied from 74.9% for prostate to 89.4% for OG. The lowest proportion of patients without a diagnosis in all three datasets was noted for breast cancer (24.2%), with the highest for lung (47.5%). For colorectal and OG, a similar proportion of

patients did not have a conformation of diagnosis in all three datasets (37.3% and 36.2% respectively).

Table 11 – Number of cancer cases recorded in each dataset. N = total number of patients with a cancer diagnosis in at least one of the datasets between 2000 – 2013.

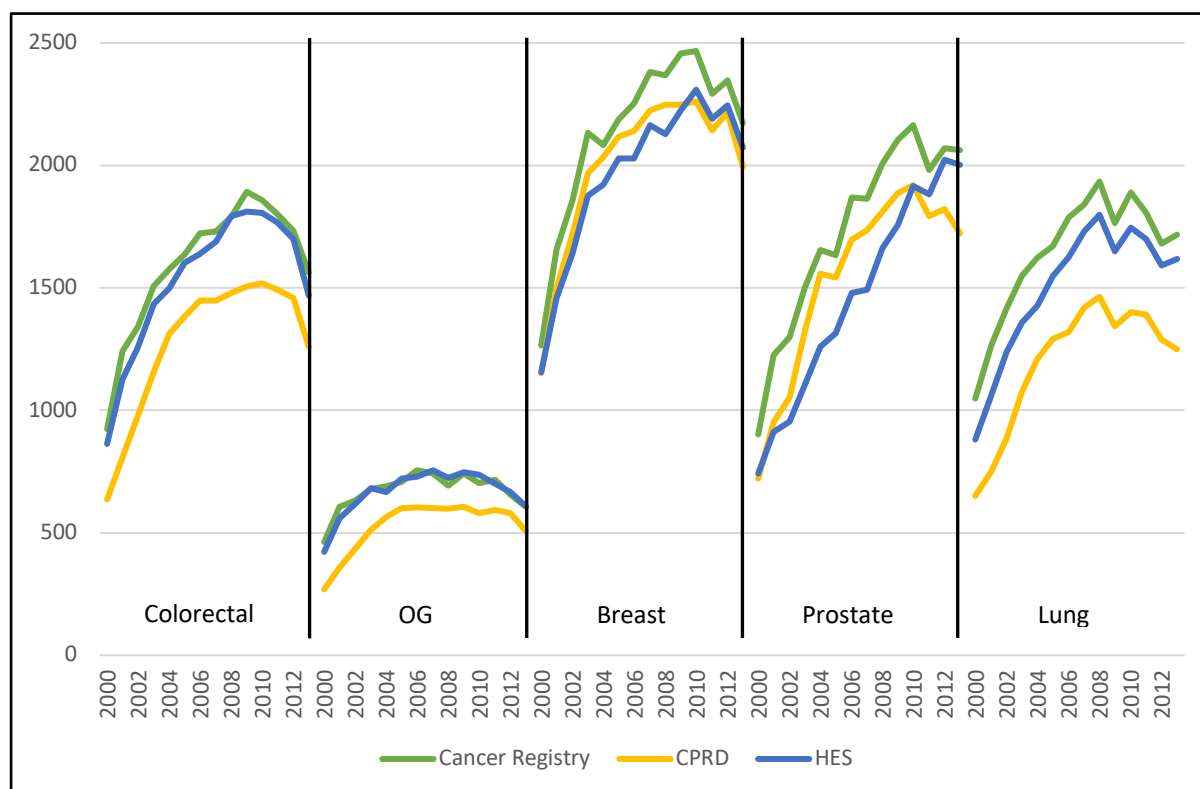
	Colorectal N = 24386		Upper GI N = 10370		Breast N = 30621		Lung N = 25581		Prostate N = 25811	
	n	%	n	%	n	%	n	%	n	%
CR	22019	90.3	9325	89.9	27538	89.9	22728	88.8	22788	88.3
CPRD	17687	72.5	7345	70.8	28912	94.4	16546	64.7	20307	78.7
HES	21131	86.7	9270	89.4	25888	84.5	20759	81.2	19329	74.9
In all three	15283	62.7	6613	63.8	23226	75.8	13436	52.5	14374	55.7

n = number of patients with a cancer diagnosis in that particular dataset between 2000 - 2013

4.1.3.2 Number of cancer diagnoses per year

For both colorectal and OG cancer, HES and the CR included a similar number of cases per year (figure 11), while CPRD had the fewest. For the remaining three cancers, the CR contained a higher number of cancers per year compared to the other two datasets. For both breast and prostate cancer, CPRD contained a higher number of patients up until 2010, after which the number were similar for breast, but the number of prostate patients in HES increased. For lung, HES and CPRD consistently showed fewer cases per year. For all cancers, the trend of cases per year for each dataset was parallel, and in particular a peak between around 2010, after which there was a fall in numbers.

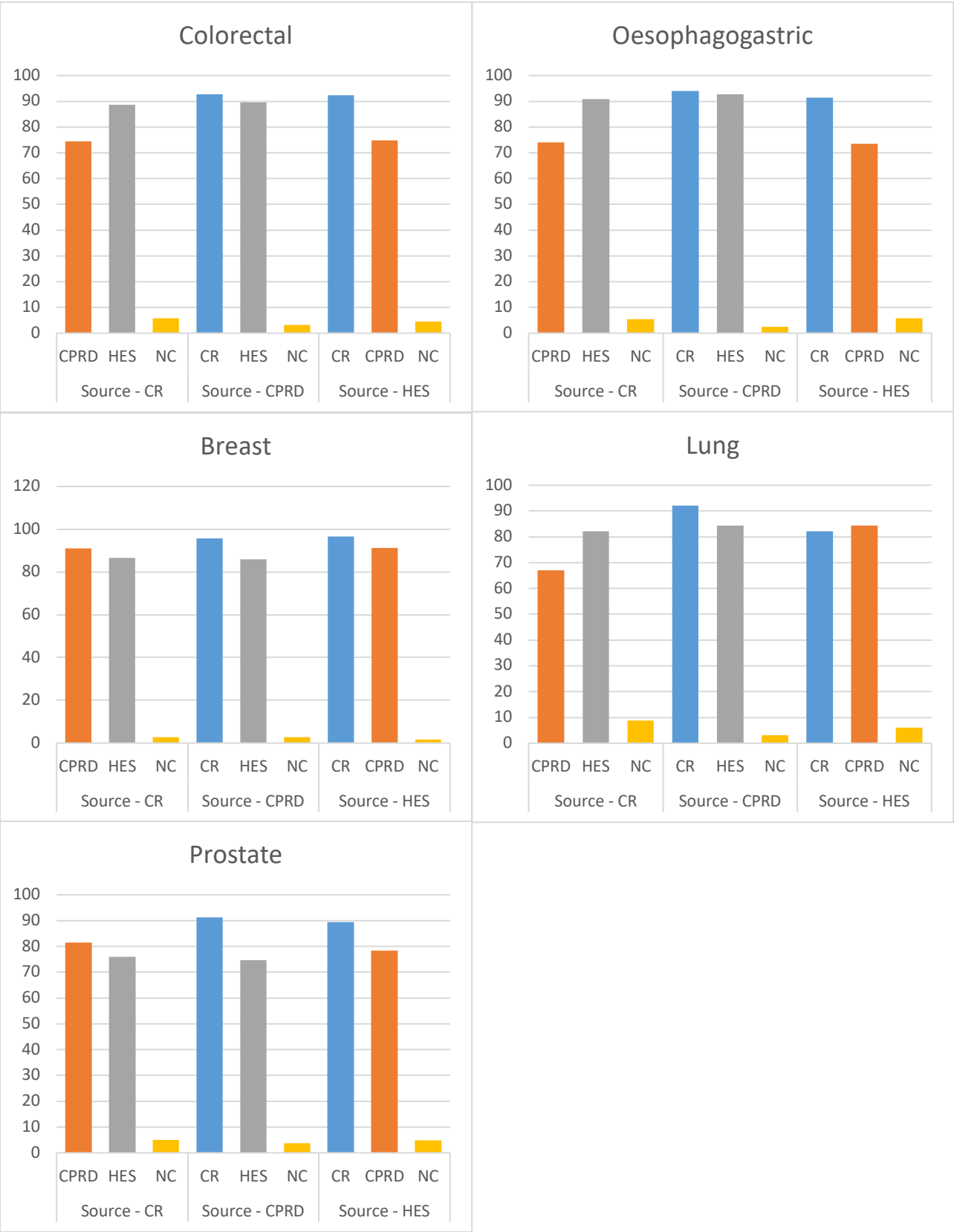
Figure 11 - Number of patients per year for each of the datasets, stratified according to cancer site



4.1.3.3 Concordance

For each cancer, over 90% of cases identified from CPRD or HES were confirmed in the cancer registry (figure 12). On the contrary, the percentage of CR cases confirmed in CPRD was relatively low for colorectal (74.5%), UGI (74.0%) and lung (67.0%) but more complete for breast (91.1%) and prostate (81.4%). The percentage of CR cases confirmed in HES was highest for colorectal (88.6%) and oesophagogastric (90.8%) cancer and breast (86.5%), but lower for lung (82.2%) and prostate (75.9%). HES and CPRD demonstrated a similar level of concordance. The CR had the highest proportion of patients without a confirmed diagnosis in either of the other two datasets for all five cancers. Only for lung was the percentage of HES cases confirmed in CPRD (84.4%) higher than the CR (82.2%).

Figure 12 - Coverage of cancer diagnoses in source dataset by the other two available dataset for each of the five cancers of interest. NC = No coverage



4.1.3.4 Difference in patient characteristics between the datasets

4.1.3.4.1 Colon cancer

There was no difference between the datasets in terms of the proportion of males or IMD (table 12). CPRD had a significantly higher proportion of younger patients than both the other datasets, while the proportion of patients over 80 was highest in the cancer registry. HES included the highest proportion of patient with a Charlson score of more than one (12.7%) than either CPRD (11.2%) or the CR (11.9%). Patients in CPRD had the highest survival after one (75.2%) or five (50.3%) years, while the CR had the lowest (69.6% and 46.1% respectively).

Table 12 - Comparison of patient demographics and survival for colorectal cancer patients depending on the source

		Cancer registry		CPRD		HES		p value
		N	%	N	%	N	%	
Gender	Male	11957 ^a	54.3	9731 ^a	55.0	11598 ^a	54.9	0.299
	Female	10062 ^a	45.7	7956 ^a	45.0	9533 ^a	45.1	
Age	Under 50	944 ^a	4.3	778 ^a	4.4	929 ^a	4.4	< 0.0005
	50 - 59	2302 ^b	10.5	2002 ^a	11.3	2277 ^{a,b}	10.8	
	60 - 69	5140 ^b	23.3	4404 ^a	24.9	5054 ^{a,b}	23.9	
	70 - 79	7122 ^a	32.3	5767 ^a	32.6	6978 ^a	33.0	
	80 and over	6511 ^c	29.6	4736 ^a	26.8	5893 ^b	27.9	
IMD	1	5257 ^a	23.9	4265 ^a	24.1	4957 ^a	23.5	0.668
	2	5490 ^a	24.9	4440 ^a	25.1	5261 ^a	24.9	
	3	4532 ^a	20.6	3676 ^a	20.8	4359 ^a	20.6	
	4	3889 ^a	17.7	3093 ^a	17.5	3817 ^a	18.1	
	5	2851 ^a	12.9	2213 ^a	12.5	2737 ^a	13.0	
Charlson score	Zero	16830 ^b	76.4	13702 ^a	77.5	15988 ^b	75.7	< 0.0005
	One	2566 ^a	11.7	2001 ^a	11.3	2461 ^a	11.6	
	More than one	2623 ^a	11.9	1984 ^a	11.2	2682 ^b	12.7	
Emergency	No	17765 ^a	80.7	14792 ^a	83.6	16270 ^a	77.0	< 0.0005
	Yes	4254 ^a	19.3	2895 ^a	16.4	4861 ^a	23.0	
Death within a year	No	15319 ^a	69.6	13294 ^b	75.2	14986 ^c	70.9	< 0.0005
	CRC related death	5404 ^a	24.5	3626 ^b	20.5	4799 ^c	22.7	
	Non-CRC death	1296 ^a	5.9	767 ^b	4.3	1346 ^c	6.4	
Death within 5 years	No	10144 ^c	46.1	8902 ^a	50.3	9967 ^b	47.2	< 0.0005
	CRC related death	9244 ^b	42.0	6886 ^a	38.9	8414 ^a	39.8	
	Non-CRC death	2631 ^c	11.9	1899 ^a	10.7	2750 ^b	13.0	

Superscripts represent significant difference in proportions between groups

4.1.3.4.2 Oesophagogastric cancer

Similar to colorectal cancer, there was no difference in gender or IMD (table 13). Overall there was no significant difference in age, although CPRD demonstrated a significantly lower proportion of patients aged 80 and over (29.4%) than either the CR (31.9%) or HES (30.8%). Again the proportion of patients with a Charlson score of zero was lowest in the HES dataset (70.4%), which also contained the highest proportion of emergency diagnoses (23.7%). The cancer registry also had the lowest percentage of patients alive at one year (30%) and five years (2.9%).

Table 13 - Comparison of patient demographics and survival for oesophagogastric cancer patients depending on the source

		Cancer Registry		CPRD		HES		p
		N	%	N	%	N	%	
Gender	Male	5984 ^a	64.2	4782 ^a	65.1	5996 ^a	64.7	0.451
	Female	3341 ^a	35.8	2563 ^a	34.9	3274 ^a	35.3	
Age	Under 50	342 ^a	3.7	276 ^a	3.8	358 ^a	3.9	0.097
	50 - 59	959 ^a	10.3	807 ^a	11.0	977 ^a	10.5	
	60 - 69	2017 ^a	21.6	1679 ^a	22.9	2068 ^a	22.3	
	70 - 79	3035 ^a	32.5	2421 ^a	33.0	3016 ^a	32.5	
	80 and over	2972 ^b	31.9	2162 ^a	29.4	2851 ^{a,b}	30.8	
IMD	1	1753 ^a	18.8	1385 ^a	18.9	1735 ^a	18.7	0.99
	2	2264 ^a	24.3	1812 ^a	24.7	2226 ^a	24.0	
	3	1868 ^a	20.1	1479 ^a	20.2	1865 ^a	20.1	
	4	1861 ^a	20.0	1438 ^a	19.6	1836 ^a	19.8	
	5	1567 ^a	16.8	1222 ^a	16.7	1595 ^a	17.2	
Charlson score	0	6730 ^a	72.2	5328 ^a	72.5	6527 ^b	70.4	0.002
	1	1221 ^a	13.1	926 ^a	12.6	1204 ^a	13.0	
	≥ 2	1374 ^a	14.7	1091 ^a	14.9	1539 ^b	16.6	
Emergency diagnosis	No	7501 ^c	80.4	6154 ^a	83.8	7069 ^b	76.3	< 0.0005
	Yes	1824 ^c	19.6	1191 ^a	16.2	2201 ^b	23.7	
Death within a year	No death	2512 ^b	30.0	2229 ^a	34.2	2541 ^b	30.9	< 0.0005
	UGI related	5242 ^a	62.7	3875 ^a	59.5	4927 ^a	59.9	
	Not UGI related	609 ^c	7.3	406 ^a	6.2	754 ^a	9.2	
Death within 5 years	No death	242 ^b	2.9	287 ^a	4.4	254 ^b	3.1	< 0.0005
	UGI related	7231 ^a	86.5	5563 ^a	85.5	6848 ^b	83.3	
	Non-UGI related	890 ^a	10.6	660 ^a	10.1	1120 ^b	13.6	

Superscripts represent significant difference in proportions between groups

4.1.3.4.3 Breast cancer

There was no significant difference in age, IMD or Charlson score between the CR and CPRD (table 14). In HES there was a lower percentage of patients aged 80 and over (13.0% vs 15.2% in CR and 14.8% in CPRD) and non-emergency diagnoses (96.7%). HES also had the lowest percentage of patients alive at one year (96.7%). A similar proportion of patients were alive at 5 years between the CR and HES, with CPRD demonstrating a significantly higher survival.

Table 14 - Comparison of patient demographics and survival for breast cancer patients depending on the source

		Cancer Registry		CPRD		HES		p
		N	%	N	%	N	%	
Age	Under 50	5183 ^a	17.9	5022 ^a	18.2	4683 ^a	18.1	< 0.0005
	50 - 59	7161 ^a	24.8	6829 ^a	24.8	6499 ^a	25.1	
	60 - 69	7262 ^a	25.1	6940 ^a	25.2	6811 ^b	26.3	
	70 - 79	4922 ^a	17.0	4680 ^a	17.0	4527 ^a	17.5	
	80 and over	4384 ^a	15.2	4067 ^a	14.8	3368 ^b	13.0	
IMD	1	7522 ^a	26.0	7242 ^a	26.3	6389 ^b	24.7	< 0.0005
	2	7353 ^a	25.4	7030 ^a	25.6	6532 ^a	25.3	
	3	5858 ^a	20.3	5584 ^a	20.3	5306 ^a	20.5	
	4	4876 ^a	16.9	4573 ^a	16.6	4555 ^b	17.6	
	5	3288 ^a	11.4	3084 ^a	11.2	3085 ^b	11.9	
Charlson score	Zero	25876 ^{a,b}	89.5	24718 ^a	89.8	23100 ^b	89.2	0.156
	One	1799 ^a	6.2	1688 ^a	6.1	1611 ^a	6.2	
	More than one	123 ^{a,b}	4.3	1132 ^a	4.1	1177 ^b	4.5	
Emergency diagnosis	No	28295 ^c	97.9	27089 ^a	98.4	25043 ^b	96.7	< 0.0005
	Yes	617 ^b	2.1	449 ^b	1.6	845 ^b	3.3	
Death within a year	No death	26971 ^c	93.3	26095 ^a	94.8	23870 ^b	92.2	< 0.0005
	Breast related	1219 ^c	4.2	885 ^a	3.2	1215 ^b	4.7	
	Not breast related	722 ^c	2.5	558 ^a	2.0	803 ^b	3.1	
Death within 5 years	No death	22898 ^b	79.2	22136 ^a	80.4	20536 ^b	79.3	0.002
	Breast related	3684 ^b	12.7	3247 ^a	11.8	3282 ^b	12.7	
	Non-breast related	2330 ^a	8.1	2155 ^a	7.8	2070 ^a	8.0	

Superscripts represent significant difference in proportions between groups

4.1.3.4.4 Prostate cancer

Age and IMD status were similar between the CR and CPRD, while HES had a higher proportion of patients over the age of 80 and lower IMD (table 15). Again HES patients had the highest Charlson score, and the highest proportion of emergency diagnoses, while CPRD patients had a lower Charlson score, and less emergency diagnoses. Patients sourced from CPRD also had the highest survival at one and five years compared to the other two datasets. Survival for the HES dataset was substantially lower at both one and 5 years (61.0% vs 81.9% in CPRD and 74.6% in CR at one year, 18.0% vs 31.3% in CPRD and 27.5% in CR at 5 years).

Table 15 - Comparison of patient demographics and survival for prostate cancer patients depending on the source

		Cancer registry		CPRD		HES		p
		N	%	N	%	N	%	
Age	Under 50	160 ^a	0.7	137 ^a	0.7	137 ^a	0.7	< 0.0005
	50 - 59	1943 ^a	8.5	1782 ^a	8.8	1608 ^a	8.3	
	60 - 69	6963 ^a	30.6	6252 ^a	30.8	5654 ^b	29.3	
	70 - 79	8618 ^a	37.8	7771 ^a	38.3	7074 ^b	36.6	
	80 and over	5104 ^c	22.4	4365 ^a	21.5	4856 ^b	25.1	
IMD	1	6265 ^a	27.5	5669 ^a	27.9	5116 ^b	26.5	< 0.0005
	2	6094 ^a	26.7	5508 ^a	27.1	5103 ^a	26.4	
	3	4681 ^a	20.5	4163 ^a	20.5	3923 ^a	20.3	
	4	3500 ^a	15.4	3097 ^a	15.3	3066 ^a	15.9	
	5	2248 ^c	9.9	1870 ^a	9.2	2121 ^b	11.0	
Charlson score	Zero	18393 ^c	80.7	16693 ^b	82.2	14981 ^b	77.5	< 0.0005
	One	2253 ^a	9.9	1964 ^b	9.7	2040 ^b	10.6	
	More than one	2142 ^c	9.4	1650 ^b	8.1	2308 ^b	11.9	
Emergency diagnosis	No	21706 ^a	95.3	19616 ^b	96.6	17412 ^c	90.1	< 0.0005
	Yes	1082 ^a	4.7	691 ^b	3.4	1917 ^c	9.9	
Death within a year	No death	7471 ^c	74.6	7000 ^a	81.9	5957 ^b	61.9	< 0.0005
	Prostate related	1412 ^c	14.1	858 ^a	10.0	1857 ^b	19.0	
	Not Prostate related	1130 ^c	11.3	691 ^a	8.1	1949 ^b	20.0	
Death within 5 years	No death	2758 ^c	27.5	2674 ^y	31.3	1759 ^b	18.0	< 0.0005
	Prostate related	3823 ^b	38.2	3056 ^a	35.7	3807 ^b	39.0	
	Non-prostate related	3432 ^a	34.3	2819 ^a	33.0	4197 ^b	43.0	

Superscripts represent significant difference in proportions between groups

4.1.3.4.5 Lung cancer

There was no significant difference in terms of gender between the three datasets (table 16). The age distribution was similar between CPRD and HES, although patients from the CR had a significantly higher proportion of patients aged 80 and over (27.4%). CPRD had a similar Charlson score profile to the CR, with HES patients demonstrating the lowest proportion of Charlson score zero. CPRD also had the highest proportion alive at both one (28.0%) and five years (4.9%), compared to the CR (21.8% and 1.8%) and HES (21.5% and 1.9%).

Table 16 - Comparison of patient demographics and survival for lung cancer patients depending on the source

		Cancer Registry		CPRD		HES		p
		N	%	N	%	N	%	
Gender	Male	13061 ^a	57.5	9442 ^a	57.1	11955 ^a	57.6	0.576
	Female	9667 ^a	42.5	7104 ^a	42.9	8804 ^a	42.4	
Age	Under 50	612 ^a	2.7	462 ^{a b}	2.8	640 ^b	3.1	
	50 - 59	2200 ^b	9.7	1726 ^a	10.4	2221 ^a	10.7	
	60 - 69	5625 ^b	24.7	4402 ^a	26.6	5513 ^a	26.6	
	70 - 79	8057 ^a	35.4	5920 ^a	35.8	7347 ^a	35.4	
	80 and over	6234 ^b	27.4	4036 ^a	24.4	5038 ^a	24.3	
IMD	1	3878 ^b	17.1	2808 ^b	17.0	3560 ^b	17.1	< 0.0005
	2	4705 ^b	20.7	3396 ^b	20.5	4314 ^a	20.8	
	3	4485 ^a	19.7	3301 ^a	20.0	4075 ^a	19.6	
	4	4844 ^b	21.3	3496 ^a	21.1	4438 ^b	21.4	
	5	4816 ^b	21.2	3545 ^b	21.4	4372 ^b	21.1	
Charlson score	Zero	14777 ^a	65.0	10837 ^a	65.5	13243 ^b	63.8	< 0.0005
	One	3761 ^a	16.5	2692 ^a	16.3	3386 ^a	16.3	
	More than one	4190 ^a	18.4	3017 ^a	18.2	4130 ^b	19.9	
Emergency diagnosis	No	17093 ^c	75.2	13086 ^a	79.1	14034 ^b	67.6	< 0.0005
	Yes	5635 ^c	24.8	3460 ^a	20.9	6725 ^b	32.4	
Death within a year	No death	4615 ^b	21.8	4266 ^a	28.0	4115 ^b	21.5	< 0.0005
	Lung related	14596 ^c	69.0	9645 ^a	63.3	12644 ^b	66.2	
	Not Lung related	1947 ^a	9.2	1336 ^a	8.8	2352 ^b	12.3	
Death within 5 years	No death	373 ^b	1.8	751 ^a	4.9	368 ^b	1.9	< 0.0005
	Lung related	18127 ^b	85.7	12461 ^a	81.7	15582 ^a	81.5	
	Non-lung related	2658 ^c	12.6	2035 ^a	13.3	3161 ^b	16.5	

Superscripts represent significant difference in proportions between groups

4.1.3.4.6 Overall comparison of datasets

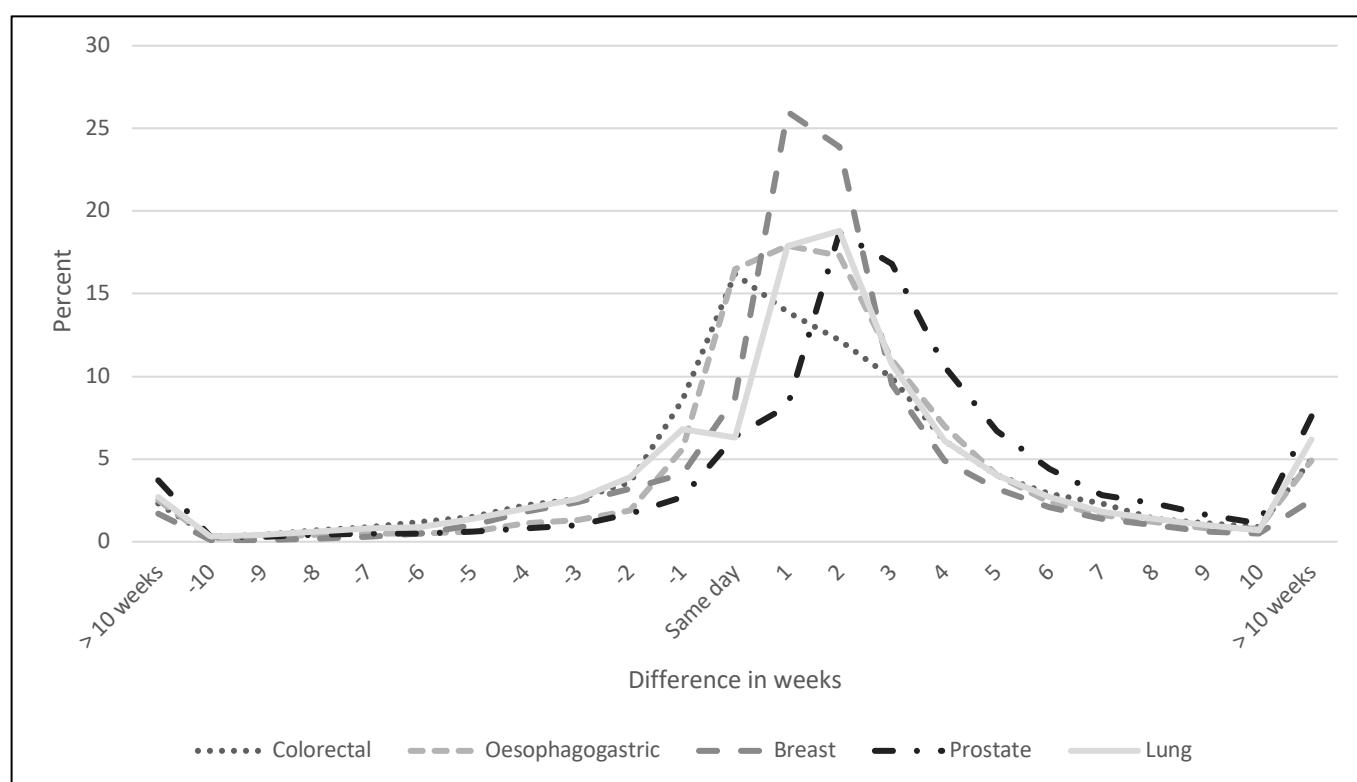
Except for breast cancer, CPRD demonstrated a significantly lower proportion of patients aged 80 and over compared to the CR. CPRD also had a significantly higher percentage of patients with a Charlson score of zero compared to the CR for colorectal (77.5% vs 76.4%) and prostate cancer (80.7% vs 82.2%), while no significant difference was seen for lung, prostate or breast cancer. For all cancers, CPRD patients demonstrated the best survival at 1 and 5 years compared to either HES or the CR. HES patients had a significantly higher Charlson score and emergency diagnoses than either CPRD or the CR for any of the five cancers, and a lower IMD level for breast, prostate and lung cancer. There was no significant difference in gender between the three datasets for the five cancers, or IMD between the CR and CPRD.

4.1.3.5 Difference in dates

4.1.3.5.1 Between CPRD and the CR

16027 (65.7%) colorectal, 6791 (65.4%) OG, 25895 (84.6%) breast, 14858 (58.1%) lung and 17488 (67.8%) prostate cancer patients had a diagnosis within a year in both CPRD and CR (figure 13). For all five cancers the majority of patients had a diagnosis in CPRD later than the CR (59.6% colorectal, 68.6% OG, 75.9% breast, 71.3% lung, 81.1% prostate) (figure 13). The largest median difference was noted for prostate cancer (16 days IQR 7 – 30), while the least difference was seen for colorectal cancer (6 days IQR 0 – 21). Colorectal and UGI cancer had the highest concordance of diagnosis date at 16.2% and 16.5% respectively.

Figure 13 - Distribution of differences in diagnosis dates between CPRD and CR.



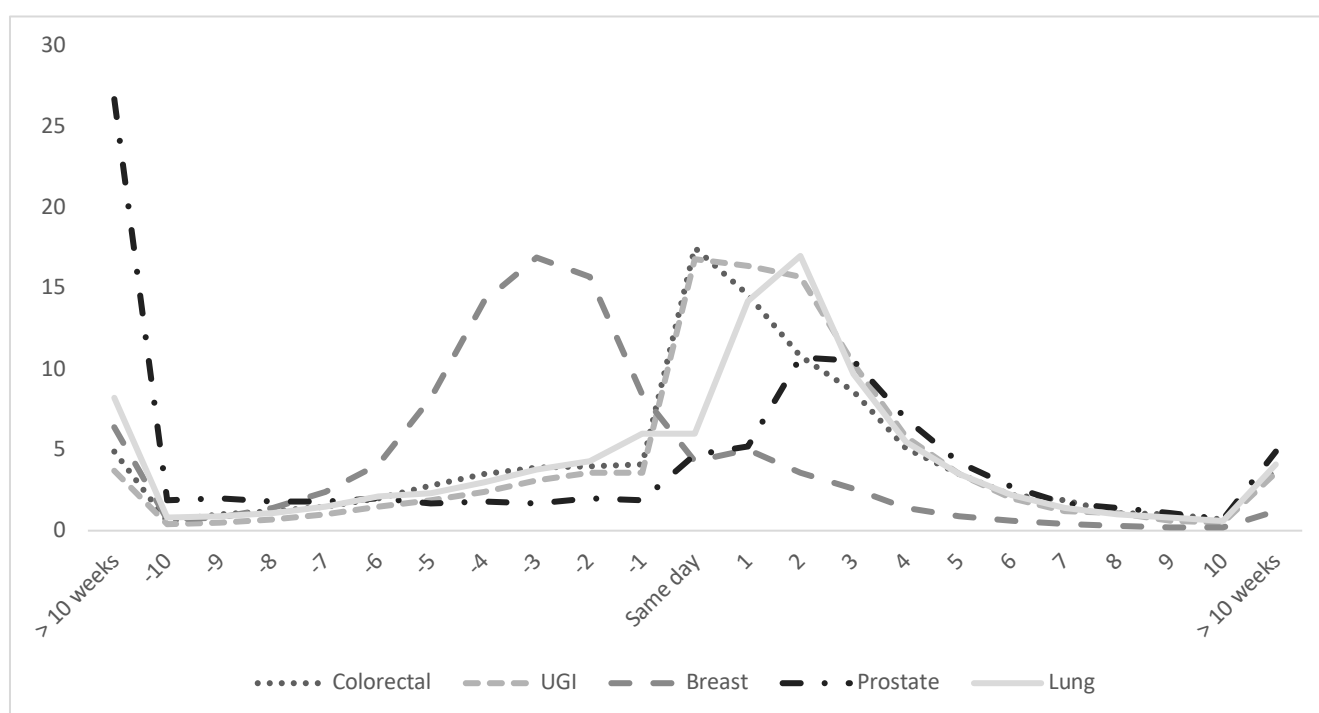
	Colorectal N = 16027		OG N = 6791		Breast N = 25895		Lung N = 14858		Prostate N = 17491	
Median days (IQR)*	6 (0 – 21)		8 (0 – 20)		7 (1 – 15)		8 (0 – 21)		16 (7 – 30)	
Same date	2592	16.2	1122	16.5	2244	8.7	935	6.3	1117	6.4
Earlier in CPRD	3879	24.2	994	14.9	4021	15.4	3324	22.4	2220	12.5
Later in CR	9556	59.6	4675	68.6	19630	75.9	10599	71.3	14154	81.1

*Positive values on the x axis and in the table indicate a diagnosis later in CPRD

4.1.3.5.2 Between CPRD and HES

15415 (63.2%) colorectal, 6688 (64.5%) OG, 22451 (73.3%) breast, 13505 (52.3%) prostate and 10986 (42.9%) lung cancer patients had a diagnosis within a year in both CPRD and HES (figure 14). The majority of colorectal, OG, prostate and lung cancer patients had a diagnosis later in CPRD compared to HES, in contrary to breast cancer for which 79.2% had a diagnosis earlier in CPRD (figure 14). The distribution of difference in diagnosis dates for prostate cancer patients demonstrated a peak at over 10 weeks earlier in CPRD and a second smaller peak at two weeks later in CPRD. For breast cancer a peak of 16.9% was seen at three weeks earlier in CPRD with second smaller peak at one week later in CPRD. The distributions for lung, UGI and colorectal were similar, with peak frequencies later in CPRD.

Figure 14 - Distribution of difference in diagnosis dates between CPRD and HES.



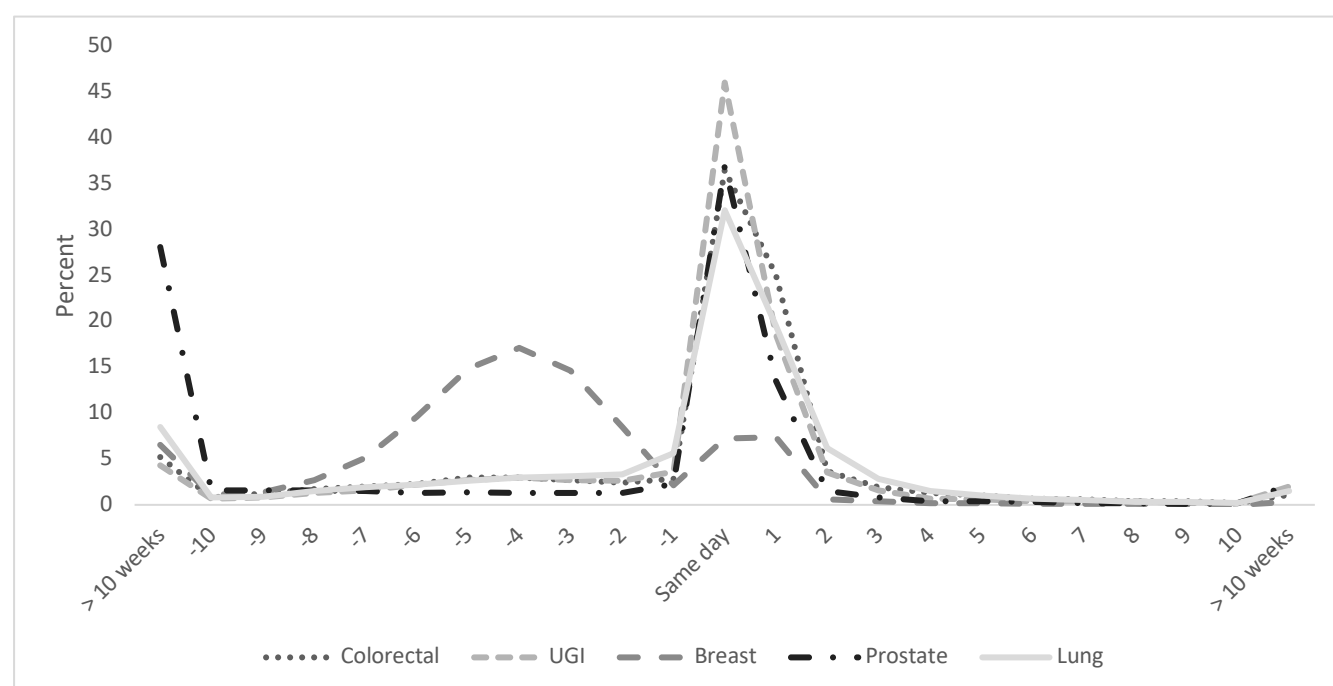
*A positive result on the x axis and in the table indicates a diagnosis later in CPRD.

	Colorectal N = 15415		OG N = 6688		Breast N = 22451		Prostate N = 13505		Lung N = 10986	
Median days (IQR)*	1 (-8 to 17)		5 (0 to 17)		-17 (-28 to -5)		6 (-13 to 17)		0 (-77 to 21)	
Same date	2698	17.5	1125	16.9	971	4.3	817	6.0	517	4.7
Earlier in CPRD	4520	29.3	1490	22.3	17803	79.2	4594	34.0	4978	45.3
Later in CPRD	8197	53.2	4073	61.1	3677	16.5	8094	60.0	5491	50.0

4.1.3.5.3 Difference in dates between the CR and HES

19266 (79.0%) colorectal, 8363 (80.6%) OG, 23972 (78.3%) breast, 18212 (70.6%) prostate and 13208 (51.6%) lung cancer patients had a diagnosis within a year in the CR and HES (figure 15). These two datasets demonstrated the highest proportion of patients with matching diagnosis dates, ranging from 46.0% for OG to 7.2% for breast (figure 15). Apart from the latter, the remaining cancers had a zero median day difference between the two datasets. Two similar peaks as in the comparison between CPRD and HES were seen again for breast and prostate cancer.

Figure 15 - Distribution of difference in diagnosis dates between CR and HES.



	Colorectal N = 19266		OG N = 8363		Breast N = 23972		Prostate N = 18212		Lung N = 13208	
Median days (IQR)**	0 (-2 to 1)		0 (-1 to 1)		-25 (-36 to -14)		0 (-14 to 2)		0 (-84 to 0)	
Same date	7027	36.5	3850	46.0	1714	7.2	5839	32.1	4876	36.9
Earlier in CR	5222	27.1	2126	25.4	20008	83.5	6069	33.3	5724	43.3
Later in CR	7017	36.4	2387	28.5	2250	9.4	6304	34.6	2608	19.7

*A positive result on the x axis and in the table indicates a diagnosis later in CR

4.1.3.6 Odds ratio for a later diagnosis record in CPRD compared to the cancer registry

There were significantly increased odds of a later diagnosis record in CPRD following an emergency diagnosis for colorectal (OR 1.48 95% CI 1.35 – 1.61), OG (OR 1.46 95% CI 1.26 – 1.70) and lung cancer (1.32 95% CI 1.20 – 1.44), while no association was seen for breast or prostate cancer (table 17). For these cancers, there was no association for a later diagnosis in CPRD compared with the CR. For both breast and prostate cancer, patients aged 70 – 79 and 80 and over were less likely than those aged 60 – 69 to have a later diagnosis in CPRD. A Charlson score of over one was associated with a later diagnosis in CPRD for colorectal (1.17 95% CI 1.06 – 1.30) and OG cancer (1.20 95% CI 1.02 – 1.40). A patient's gender demonstrated no association with the timing of a cancer diagnosis.

Table 17 - Odds ratio for a cancer diagnosis recorded later in CPRD

		Colon		OG		Breast		Lung		Prostate	
		OR	p	OR	p	OR	p	OR	p	OR	p
Age	Under 50	1.15 (0.97 – 1.36)	0.107	1.34 (0.99 - 1.82)	0.063	0.95 (0.87 – 1.04)	0.241	0.93 (0.74 - 1.16)	0.500	2.20 (1.15 – 4.22)	0.017
	50 - 59	0.97 (0.86 - 1.08)	0.552	0.89 (0.74 – 1.08)	0.232	0.93 (0.86 - 1.01)	0.105	1.00 (0.88 - 1.14)	0.971	0.88 (0.75 - 1.02)	0.081
	60 - 69	ref									
	70 - 79	0.95 (0.88 - 1.04)	0.255	0.96 (0.84 - 1.11)	0.601	0.75 (0.68 – 0.81)	< 0.0005	1.02 (0.930 - 1.12)	0.695	0.86 (0.78 – 0.94)	0.001
	≥ 80	0.93 (0.85 - 1.02)	0.100	0.96 (0.83 -1.21)	0.559	0.65 (0.59 – 0.72)	< 0.0005	1.04 (0.94 - 1.15)	0.505	0.55 (0.50 – 0.62)	< 0.0005
Gender	Female	ref									
	Male	0.97 (0.89 - 1.02)	0.129	0.99 (0.89 – 1.11)	0.908	-		1.05 (0.97 – 1.13)	0.220	-	-
IMD	1	ref									
	2	0.96 (0.88 - 1.05)	0.393	0.94 (0.80 - 1.10)	0.453	0.92 (0.85 – 0.99)	0.036	0.91 (0.80 - 1.02)	0.100	0.97 (0.87 - 1.08)	0.555
	3	0.95 (0.87 - 1.05)	0.310	0.99 (0.84 - 1.17)	0.893	0.89 (0.82 – 0.97)	< 0.0005	0.86 (0.77 – 0.97)	0.015	0.88 (0.78 – 0.98)	0.018
	4	1.00 (0.90 - 1.11)	0.954	0.89 (0.75 - 1.04)	0.146	0.96 (0.88 - 1.05)	0.368	0.96 (0.86 - 1.08)	0.537	0.80 (0.71 – 0.90)	< 0.0005
	5	1.02 (0.91 - 1.14)	0.751	1.14 (0.96 - 1.36)	0.148	0.81 (0.74 – 0.90)	< 0.0005	0.92 (0.82 - 1.04)	0.185	0.91 (0.78 - 1.05)	0.173
Charlson score	Zero	ref									
	One	1.05 (0.96 - 1.16)	0.306	1.18 (1.01 - 1.37)	0.040	1.07 (0.95 - 1.20)	0.297	1.02 (0.93 – 1.13)	0.634	1.07 (0.95 - 1.22)	0.218
	Over one	1.17 (1.06 – 1.30)	< 0.0005	1.20 (1.02 – 1.40)	0.027	1.00 (0.87 - 1.16)	0.992	1.04 (0.97 - 1.13)	0.220	1.08 (0.94 - 1.24)	0.299
Emergency	No	ref									
	Yes	1.48 (1.35 – 1.61)	< 0.0005	1.46 (1.26 – 1.70)	< 0.0005	1.13 (0.90 - 1.43)	0.293	1.32 (1.20 – 1.44)	< 0.0005	0.82 (0.68 – 1.00)	0.050

4.1.4 Discussion

4.1.4.1 Summary

This is the first comprehensive study of cancer diagnosis recording between three of the largest national datasets available in the UK. I described the discrepancies in diagnosis dates and differences in patient demographics between CPRD, CR and HES. I demonstrated that patient characteristics changed according to the data source as well as the timing of diagnosis. Overall the cancer registry confirmed the highest proportion of patients, while CPRD had the lowest coverage. The lowest median difference in diagnosis dates, together with highest proportion of patients with a matching diagnosis date, was noted between CR and HES for all cancers except breast. For the later the CR and CPRD demonstrated the highest concordance in dates.

Delays in diagnosis and treatment have been identified as potential causes leading to the worse outcome following cancer diagnosis in the UK compared to Northern and European countries²⁷. However due to the different methods of data collection and definitions of the diagnosis date, there are mixed conclusions about whether delays are associated with a worse prognosis. The Aarhus statement regarding the patient pathway has stated the diagnosis date should correspond to the first histological confirmation¹. Routinely collected national data, as found in the cancer registry, HES and CPRD, provides a large cohort for epidemiological studies to investigate such delays¹²¹. As the information has not been obtained especially for the purpose of a particular study, these datasets have an advantage over bespoke data by saving both time and cost, while excluding recall bias¹²⁹. However the robustness of these studies depends on the accuracy of the dataset being used. If the dataset has a systematic deficiency in the process of data collection, then there is an element of bias that will be inevitably introduced into the study. The first studies describing the relatively poor outcome in the UK was based on cancer registry data that had a relatively low coverage of cancer cases¹⁴⁰. Since the introduction of international standards for data collection, the cancer registry has improved case ascertainment in recent years¹³⁰. It is now used as an important source for international comparison of hospital care and can be considered the gold standard relative to HES and CPRD.

HES has been shown to be accurate when compared to hospital clinical notes¹²⁹. However our study has shown that HES-only studies are likely to over represent older patient and those diagnosed through an emergency route, resulting in the worse survival noted. On the contrary, CPRD-only studies compared to CR will under-represent these groups. The relatively better survival in CPRD patients may also reflect a lack of cancer diagnosis for patients who die shortly after the first cancer diagnosis in hospital.

Our findings confirm previous studies that although a high proportion of CPRD diagnoses can be verified with the cancer registry, a relatively low percentage of cases found in the cancer registry are found in CPRD^{127,141}. This is particularly an issue with colorectal and OG, both of which are associated with a higher emergency route to diagnosis than the other cancers. The similar level of coverage of CPRD by HES confirms this.

The discrepancy in dates found in this study correlate to the different routes of diagnosis for each cancer. For both prostate and breast cancer two peaks were identified in the comparison between HES and the other two datasets. The first peak corresponds to patients diagnosed through outpatient clinics, recorded in CPRD or the CR but not in the inpatient HES dataset used in this study. For prostate cancer, there may be a prolonged period between neo-adjuvant treatment and resection, explaining why the peak occurs at over 10 weeks earlier in CPRD compared to HES, compared to just 4 weeks for breast cancer. A secondary peak indicating a diagnosis in HES earlier than CPRD identifies patients who either underwent a resection or were diagnosed following an inpatient procedure. The relatively low proportion of CPRD cases covered in HES for these two cancers may be explained by the availability of hormonal treatment as first-line.

The relatively high concordance in diagnosis dates between HES and CR for the remaining three cancers indicates use of an endoscopic examination leading to a tissue diagnosis. Inpatient HES has been previously used for studies based on endoscopic procedures¹⁴². As the GP is likely to be informed of the diagnosis through a letter from the hospital, the diagnosis in CPRD is found later in the majority of patients. Through regression modelling I also found colorectal, OG and lung cancer patients diagnosed through an emergency were more likely to have a diagnosis recorded in CPRD later than the CR. Breast and prostate

cancer patients aged over 69 were less likely to have a later diagnosis in CPRD, perhaps reflecting the presumptive diagnosis of cancer following finding of a breast mass or raised Prostatic Specific Antigen in primary care, or the diagnosis date in the cancer registry was changed after further histological data was provided following a resection.

4.1.4.2 Comparison with previous studies

By using linkage between the three datasets used in our study, Dregan et al.¹²⁷ also found the percentage of cases in CPRD recorded in the cancer registry was high and the majority of patients had a diagnosis in the cancer registry earlier than CPRD. However they did not investigate comparisons to HES, nor the difference in demographics between datasets. Boggon et al.¹²⁸ demonstrated the percentage of cases in CPRD confirmed in CR was not as high for breast and prostate cancer as I found in our study, which may be because their cohort was originally extracted for the purposes of a diabetes related study. They described the differences in patient characteristics for all the cancers combined. By presenting data for each of the five most common cancers separately, I have found differences in characteristics which would have been missed if amalgamated together.

From our findings I conclude studies based solely on CPRD will incorrectly prolong the diagnostic interval and shorten survival, as the latter is often calculated from the diagnostic date. This will falsely add weight to an association between delays and prognosis. Topping et al.⁷⁸ found a significant difference in the diagnostic interval between cohorts depending if the source was a GP questionnaire or patients record. My findings suggest the date of diagnosis should be defined by the cancer registry when possible, and the study population should be identified from this dataset to avoid exclusion of patients of certain characteristics.

4.1.4.3 Strengths and limitations

Patients were selected solely on the basis of a cancer diagnosis in at least one of the linked datasets. Previous studies comparing coverage have used cohorts selected for other conditions. As a result I was able to include a high number of patients to find differences in demographics. As dates in all three datasets were provided in a day, month and year format,

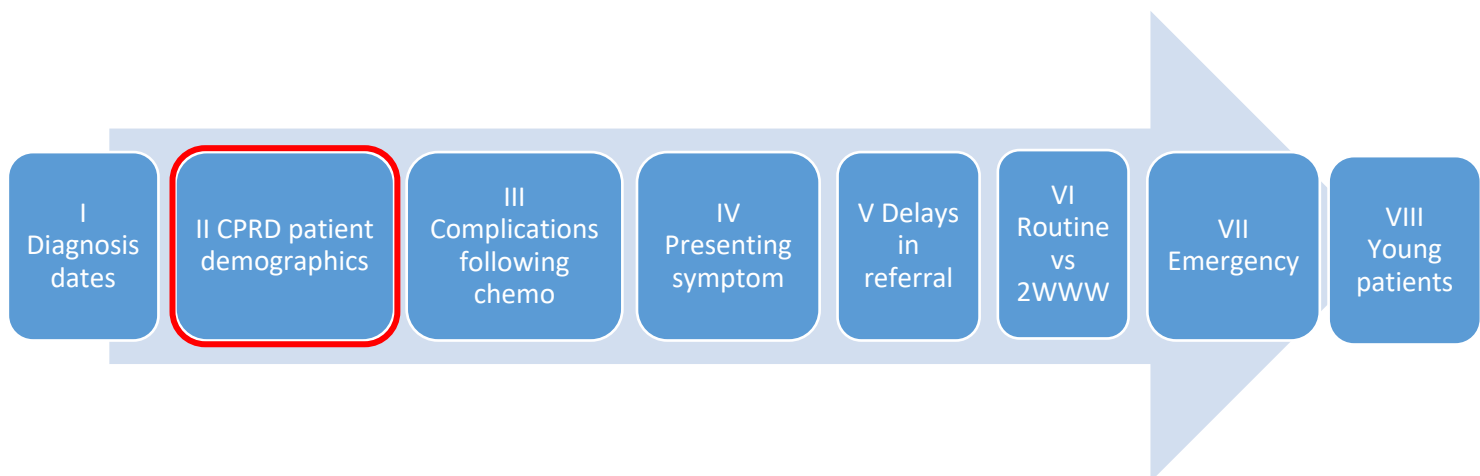
I have made accurate comparisons.

I am unable to determine whether the cancer diagnosis recorded in CPRD by the GP was a provisional diagnosis based on the clinical findings at the time of presentation. My aim was to confirm the coverage based on READ codes which represent a cancer diagnosis. Further data regarding the diagnosis of cancer may be found within the free text¹²⁴, so the coverage of CPRD is likely to be more complete than presented here. However the free text is no longer available to researchers. I would expect no difference in patient demographics dependent on whether the diagnosis was recorded as free text or a READ code. Therefore any differences in patient demographics I found between datasets are likely to be true differences rather than a result of missing data.

4.1.4.4 Conclusion related to thesis

- Epidemiological studies require robust data to allow the findings to be translated to the general population
- Cancer population should be identified from the cancer registry
- Diagnosis date should be determined from the cancer registry

End of Study



4.2 Study II - Comparison of cancer registry patients from CPRD with national published data from NCIN

The previous chapter determined that when using ‘Set 12’, as in this thesis, the patients should ideally be determined using the cancer registry rather than CPRD. However the previous study cannot conclude that the patients identified in the cancer registry from ‘Set 12’ are representative of the national cancer population.

The hypothesis of this study states:

- **Patients extracted from CPRD are representative of the national cancer population**

The aims of this study are to answer the following question:

- **Are patients extracted from the CPRD linked cancer registry similar in terms of demographics, stage and route of diagnosis to the wider national cancer population?**

4.2.1 Introduction

The National Cancer Intelligence Network (NCIN) provide yearly updates regarding the demographics and stage of colorectal cancer diagnosis¹⁵, amongst other cancers. Data from NCIN determines health policy. This chapter compares the patient demographics, route of diagnosis and stage between colorectal cancer patients from the cancer registry in the linked CPRD dataset to the national data provided by NCIN. If the two datasets are not comparable, then the conclusions regarding delays from this thesis will not be valid to the wider population.

4.2.2 Methods

The number of cases per year of colorectal cancer in the linked cancer registry between 2000 and 2013 inclusive, and their stage data, is described. The proportion of national cases found in CPRD between 2006 and 2013 inclusive is demonstrated. The latter is restricted to 2006 onwards as national data before this is not accessible.

The gender, age (as categorical data of under 50, followed by decades up to 80 and over), socioeconomic status based on Index of Multilevel Deprivation (IMD), route of diagnosis (based on Ellis-Brookes et al.²⁹ algorithm) and stage are compared between CPRD linked cancer registry patients and NCIN data.

4.3.3 Results

22019 patients had a diagnosis of colorectal cancer (ICD C18 – 20) in the linked cancer registry between 2000 – 2013 inclusive. The number of cases rose from 897 in 2000 to a peak of 1879 in 2009, followed by a fall to 1459 in 2013 (figure 16). The proportion of national cases found in the CPRD linked registry reaches a peak of 5.6% in 2009, with a fall to 4.5% in 2013 (Figure 17). Overall 5.2% of new cancer diagnoses in England between 2006 – 13 are confirmed in the CPRD linked registry.

Figure 16 – Number of colorectal cancer cases per year in the linked CPRD cancer registry

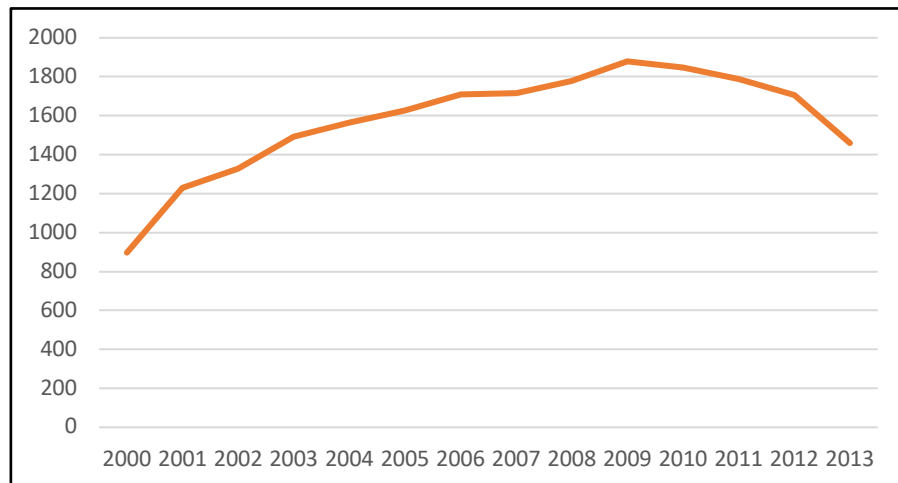
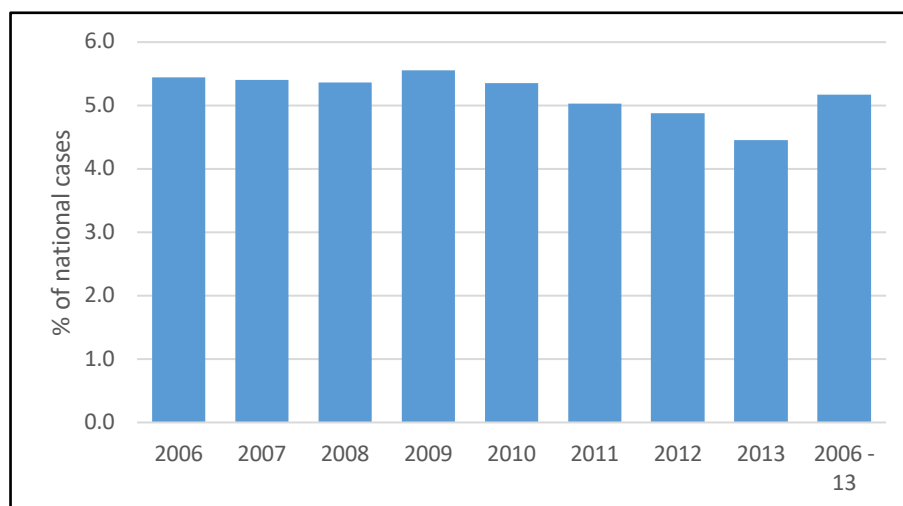
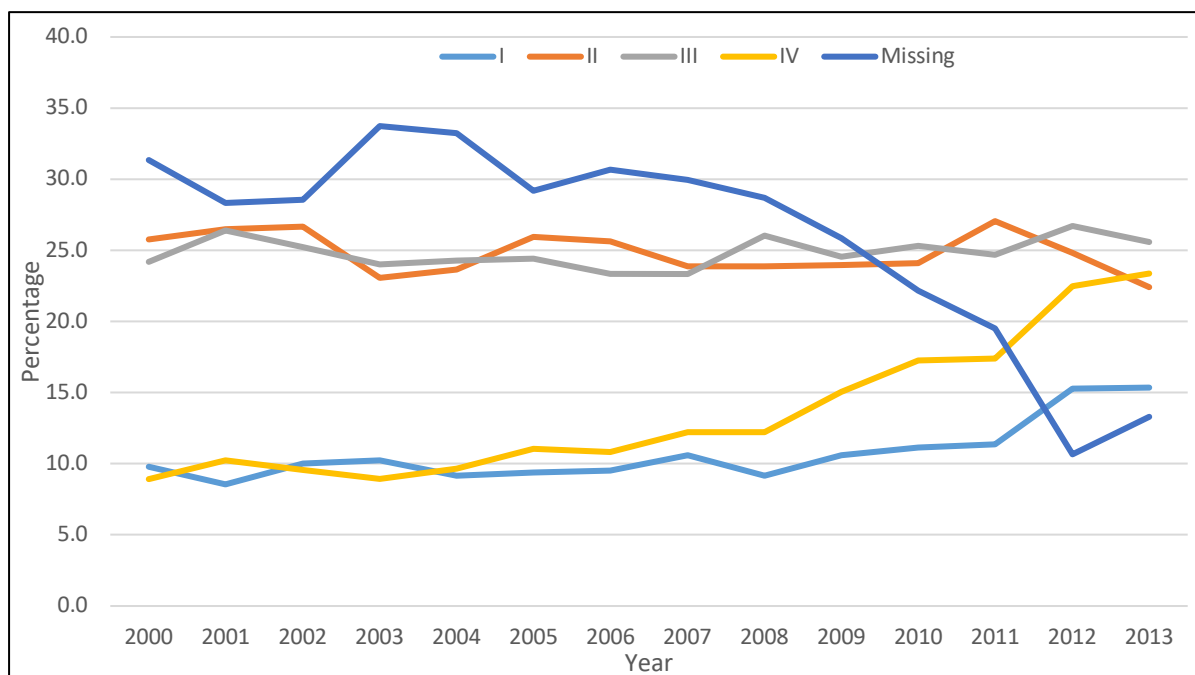


Figure 17 – Percentage of national cases in England found within the CPRD linked cancer registry



From 2000 to 2013, there is a decrease in the proportion of missing stage data from 31.3% in 2000 to a low of 10.7% in 2012 (figure 18). Stage II and III cancer remained relatively stable.

Figure 18 – Stage according to year for all colorectal cancer cases in the CPRD linked registry 2000 - 13



The proportion of patients of each category within gender, age and IMD are similar, and are within the confidence interval of the individual datasets (figure 19). Also the difference in the route of diagnosis is negligible between the datasets for cancer diagnosed between 2006 – 13 inclusive. 30% of CPRD cases were diagnosed via the TWW route, with 24% as an emergency (figure 20). Analysis of stage data between 2012-13 demonstrates equally comparable stage data between the datasets (figure 21).

The difference in numbers and cancer coverage over the years does not preclude the use of CPRD, as long as the concordance of patient demographics and stage is maintained, which adds justification to the process of cleaning the stage data described in the methods section.

Figure 19 – Patient demographics between CPRD and the national cancer population in England

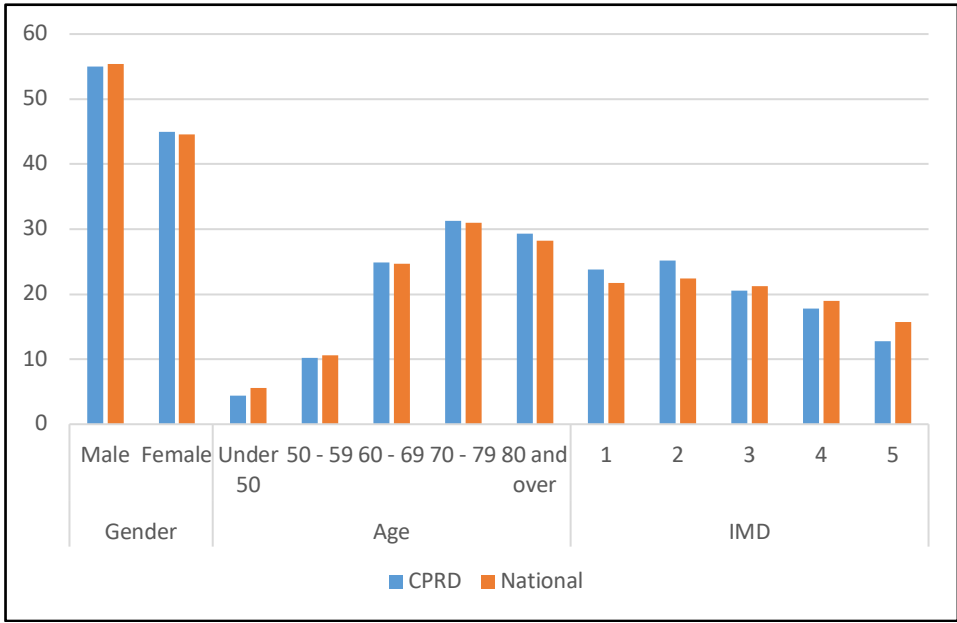


Figure 20 – Comparison of the route of diagnosis between CPRD and national cancer registry for colorectal cancer diagnoses between 2006 – 13

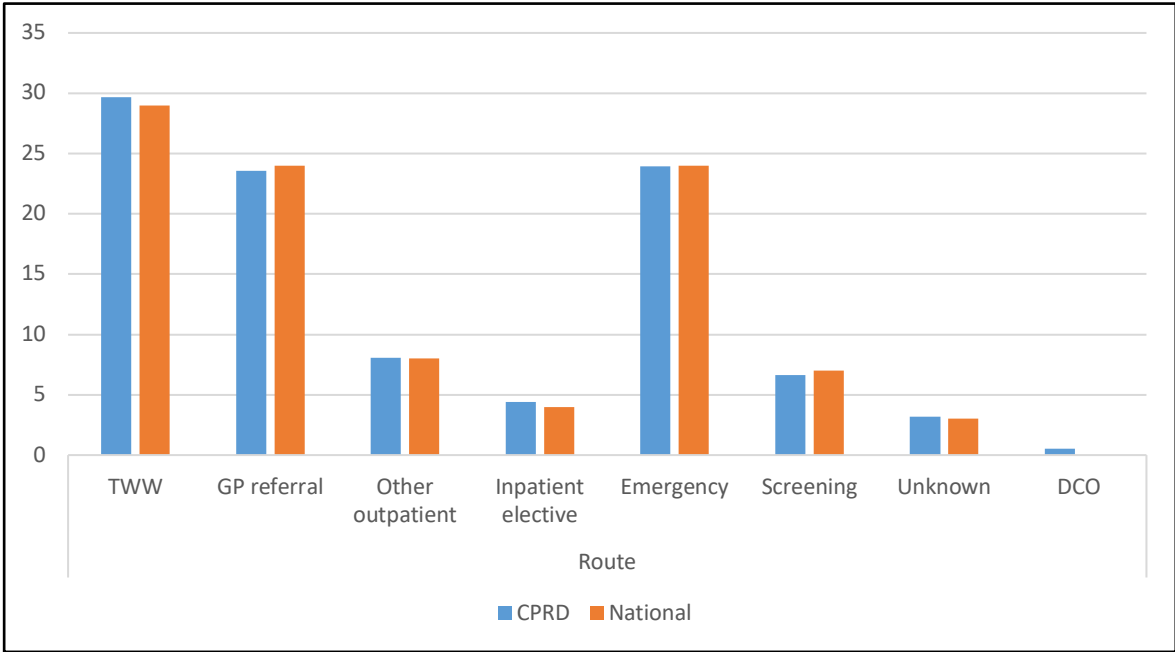
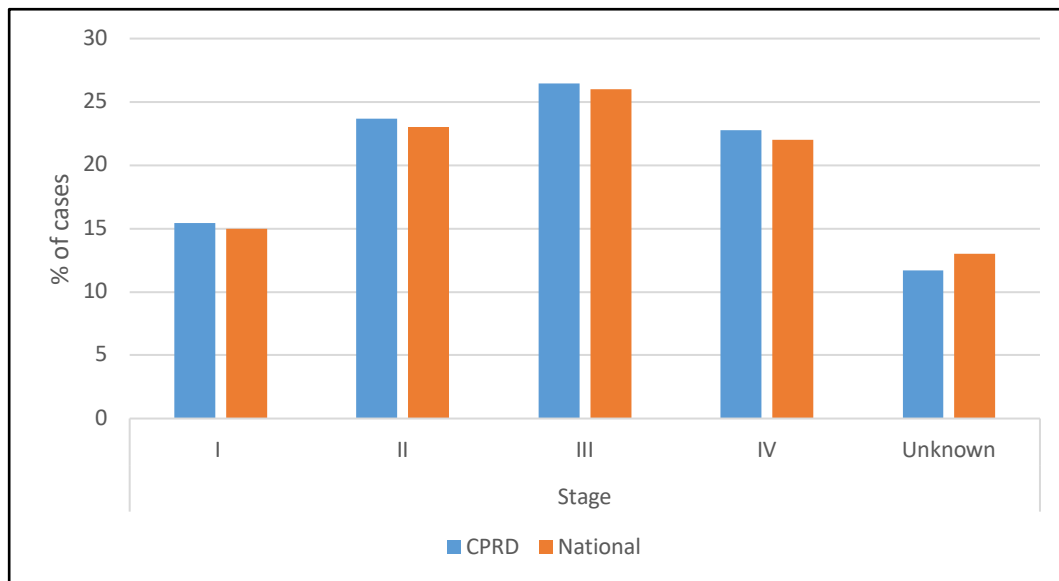


Figure 21 – Proportion of cases according to stage, compared between CPRD and the national cancer registry of England 2012-13



4.2.4 Discussion

This study demonstrates the colorectal cancer patients selected from the cancer registry in the linked CPRD, which will be used in this thesis, is representative of the national cancer registry.

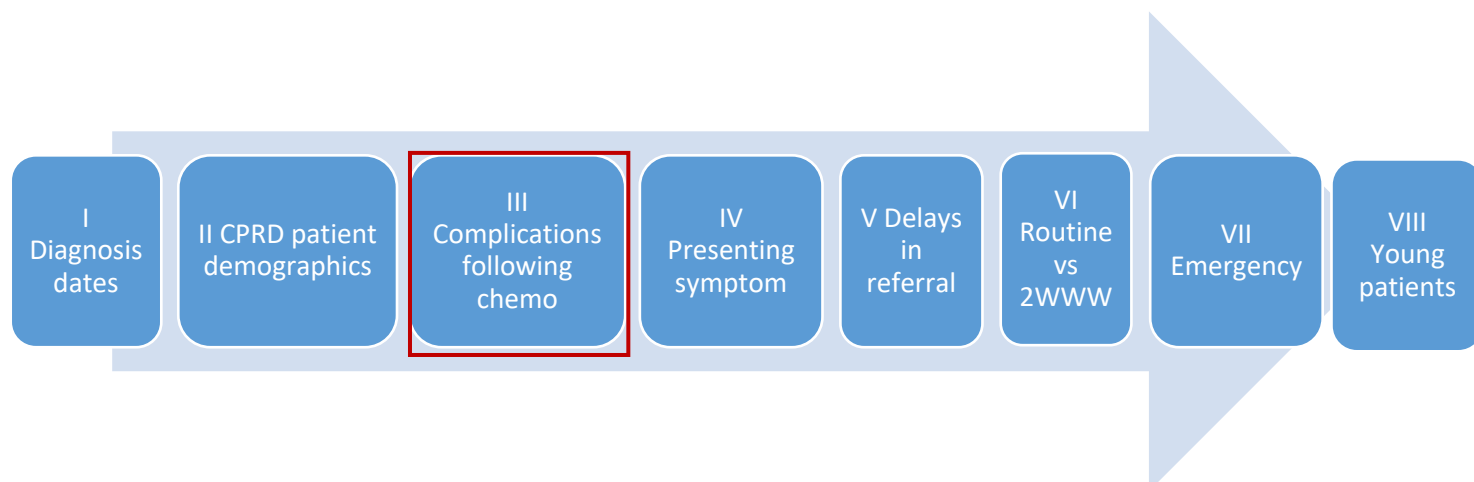
The initial rise in cases between 2000 to 2009 is likely due to the improved coverage of the larger CPRD dataset and the increase in the proportion of GP practices who have given permission to allow the patient data to be linked to the cancer registry. There has also been an increase in the completeness for the cancer registry, although this is unlikely to explain such a sharp increase. The fall post 2009 is due to the time of data collection, as there is a lag before cases are registered with the Cancer registry. As shown in figure 16, with a more up to date dataset I would expect the number of cancer cases and coverage to increase, until a slight fall in the last couple of years.

The notable decrease in missing stage data is testament to the introduction of national standards and procedures in 2006 to improve the quality of registry data. However as the proportion of missing stage decreased, stage IV and I increased. The latter increase is likely due to screening, while the majority of missing stage are likely to be stage IV, as survival after diagnosis for patients with missing stage is poor¹⁶.

4.2.4.1 Conclusion and relevance to thesis

CPRD linked cancer registry is similar to the cancer registry of England in terms of age, gender, IMD, route of diagnosis and stage. Findings from this thesis should be applicable to the wider cancer population.

End of Study



4.3 Study III - Complications following colonic resection are associated with a delay in adjuvant chemotherapy treatment which reduces overall survival

The two previous studies have confirmed patients for observational studies should ideally be identified from the cancer registry linked to CPRD rather than CPRD alone to avoid selection bias and provide accurate diagnosis dates. In addition, such patients are representative of the national cancer population. However to determine delays and their impact on prognosis, it is also necessary to determine if using linked CPRD data is a valid alternative to stand-alone HES or patient records. In the study below I compare the outcomes of delays in treatment for cancer patients to studies using HES and patient records. The additional benefit of such a study is to ensure my understanding of the linked HES and ONS datasets and regression models, before embarking on more complex analysis.

The hypotheses of this study are:

- **Complication following a resection for colon cancer is associated with a delay in adjuvant chemotherapy and worse prognosis**
- **CPRD provides data missing in HES that describes the cause of delays in the patient pathway**
- **Results from CPRD regarding delays are comparable to other national datasets**

The aims of this study are to answer the following questions:

- **What are the patient and operative factors associated with a delay in adjuvant chemotherapy beyond eight weeks?**

4.3.1 Introduction

Adjuvant chemotherapy following resection for stage III colon cancers has been shown to improve prognosis, with an absolute increase of 7% in overall 5-year survival¹⁴³. Although there are no official guidelines regarding the timing of adjuvant chemotherapy following resection, multiple studies have found a delay beyond eight weeks is associated with a reduction in the benefit of treatment^{144–146}.

Recent studies have suggested complications following resection may lead to a delay in adjuvant chemotherapy¹⁴⁷. However these studies have concentrated on hospital notes, and have not considered complications post-discharge such as venothromboembolic events or wound infections. These are often managed in primary care and have the potential to delay treatment further. In addition, no study has investigated an association between complications and a delay in adjuvant chemotherapy while accounting for the urgency of admission that led to the resection, co-morbidities and the type of surgical approach. It is unclear whether the impact of a delay in adjuvant chemotherapy on overall survival is maintained after taking into account these factors.

4.3.2 Method

4.3.2.1 Patient selection

Patients who had undergone a colon cancer resection were identified from the linked HES dataset as having a procedure based on OPCS coding (see appendix vii) and an ICD-10 coded diagnosis of colon cancer in the same admission between 1st April 2000 and 1st April 2011. From this population the study cohort was selected as those with a HES episode for chemotherapy (see appendix viii) within 16 weeks of the resection. A cut-off of 16 weeks has been previously been used as it is unlikely such late AC treatment is for curative intent following resection¹⁴⁸. Patients were excluded if they were under 18 years of age, had an ICD code indicating metastatic disease (see appendix ix) recorded within 28 days of the resection, or an alternative cancer diagnosis within three years before the resection. The individual patient records are linked to the Office for National Statistics (ONS) cause of mortality. Therefore I could describe the date of surgery, the start of adjuvant chemotherapy, the type of admission that led to the resection and date of death.

4.3.2.2 Study variables

If the first chemotherapy episode was recorded within eight weeks of the resection, the patient was identified as having received adjuvant chemotherapy early. The Chi-Square test was used to test for differences between the early and late treatment group for age (under 60, 60 to 75, over 75), gender, smoking history, Charlson score (zero, one or more), year of procedure (2000 to 2003, 2004 to 2007, 2008 to 2011) surgical access (laparoscopic or open), admission type for resection (emergency or elective), stoma at resection and a complication following a resection.

Charlson score was calculated from three years of HES data preceding the resection. As all included patients only had colon cancer with no metastases by definition, the Charlson scoring did not include cancer diagnoses. A laparoscopic resection was defined by the presence of a specific OPCS at the time of the resection. Conversion to open was considered an open operation rather than laparoscopic. Patients were identified as receiving a stoma if they had a Hartmann's procedure or an OPCS code indicating a stoma. Complications from HES (re-operations (see Appendix), acute renal failure, myocardial infarction, a cerebrovascular event and treatment for bowel obstruction without an operation) were identified in the interval between the resection and start of adjuvant chemotherapy, while complications from CPRD were considered if they occurred following discharge after the resection. Both CPRD and HES were used to identify pneumonia, a venothromboembolic event and wound complications treated without an operation. All complications were considered equally in the regression models, so patients either had zero or at least one complication.

4.3.2.3 Statistical analysis

The above variables were included in a binary logistic regression model to identify the factors associated with a delay in initiating adjuvant chemotherapy at eight weeks or later. Clinically relevant two-way interactions were entered individually with the main effects model to determine whether they produced a significant change. All such interactions were then entered together into the model and removed step-wise if found to be non-significant.

The only significant interaction included in the model was between the type of admission and surgical access.

A Cox proportional hazards model was used to identify an association between initiating adjuvant chemotherapy after eight weeks or more with all-cause mortality, while taking into consideration the aforementioned patient and surgery variables. Survival was calculated from the start of adjuvant chemotherapy to the date of death as noted in the ONS dataset. Patients were censored if no date of death was noted in the ONS data. Smoking history and a delay in adjuvant treatment was the only significant interaction and was included in the full model. Kaplan Meier curves described the difference in survival function between the early and late adjuvant chemotherapy groups.

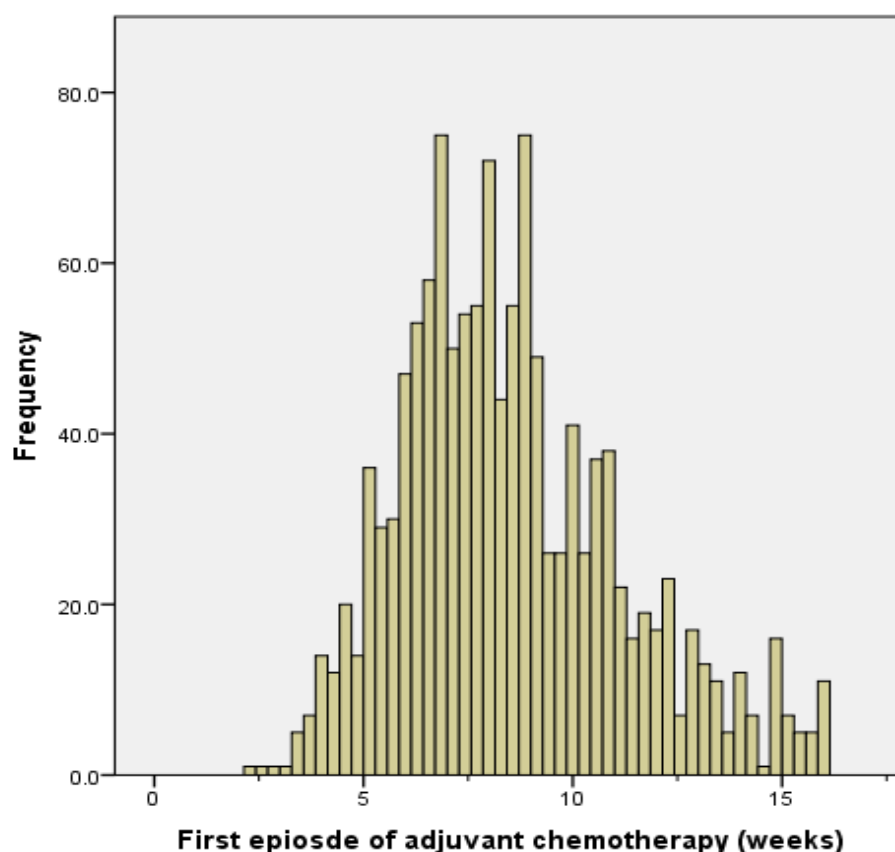
Robust standard errors were constructed that accounted for heteroscedasticity. Patients were excluded if there was missing demographic data or year of operation.

4.3.3 Results

In total 7501 patients were identified from the linked HES database as having undergone a colon cancer resection. Of these 1266 (16.9%) underwent adjuvant chemotherapy within sixteen weeks. Up until 2005 the proportion of patients identified as having undergone adjuvant chemotherapy was similar at around 14%. From 2006 there is an increase to a maximum of 23.3% in 2008 (n=155).

The median time to the first adjuvant treatment was 8.1 weeks (Interquartile range 6.7 to 10.1). The earliest was at 2.3 weeks (figure 22), with the latest as per the criterion of 16 weeks. 598 patients (47.2%) started their adjuvant chemotherapy within 8 weeks of the resection.

Figure 22 - Frequency distribution of patients according to first adjuvant chemotherapy episode



4.3.3.1 Summary statistics

Although there was no overall significant difference between the different age groups in terms of first episode of adjuvant chemotherapy, a higher proportion of patients over the age of 75 ($n = 124$, 59.9%) experienced a delay in treatment compared to those who were under 60 ($n=181$ 50.1%) and aged 60 to 75 ($n=363$ 52.0%) (table 18). A significantly higher percentage of patients receiving AC late had a stoma at the time of the resection ($n=116$ 65.5% $p<0.005$). There was no difference found between the two groups in terms of gender, year of resection or Charlson score. Also, there was no difference in the type of surgical access between the two groups (laparoscopy $n=108$, 46.8% in the early group vs $n=123$, 53.2% in the late group, $p=0.871$) nor emergency admission ($n=156$ 44.6% vs $n = 194$ 55.4% respectively, $p=0.241$).

Table 18 Summary statistics comparing patients with first episode adjuvant chemotherapy before and after eight weeks n (%)

		Total (%)	Less than 8 weeks N = 598 (47.2)	8 weeks or more N = 668 (52.8)	p value
Age	Under 60	361 (28.5)	180 (49.9)	181 (50.1)	0.067
	60 to 75	698 (55.1)	335 (48.0)	363 (52.0)	
	Over 75	207 (16.4)	83 (40.1)	124 (59.9)	
Gender	Male	693 (54.7)	337 (48.6)	356 (51.4)	0.275
	Female	573 (45.3)	261 (45.5)	312 (54.5)	
Year of resection	2000 to 2003	307 (24.2)	157 (51.1)	150 (48.9)	0.132
	2004 to 2007	386 (30.5)	187 (48.4)	199 (51.6)	
	2008 to 2001	573 (45.3)	80 (44.3)	88 (55.7)	
Stoma at resection	No Stoma	1089 (86.0)	537 (49.3)	552 (50.7)	< 0.005
	Stoma	177 (14.0)	61 (34.5)	116 (65.5)	
Admission type	Elective	916 (72.4)	442 (48.3)	474 (51.7)	0.241
	Emergency	350 (27.6)	156 (44.6)	194 (55.4)	
Surgical access	Open	1035 (81.8)	490 (47.3)	545 (52.7)	0.871
	Laparoscopic	231 (18.2)	108 (46.8)	123 (53.2)	
Charlson score	Zero	975 (77.0)	466 (47.8)	509 (52.2)	0.465
	I or more	291 (23.0)	132 (45.4)	157 (54.6)	

4.3.3.2 Complications

Through either HES or CPRD, 68 patients (5.4%) were found to have suffered at least one systemic complication: 21 (3.5%) patients in the early group compared to 47 (7.0%) in the late group ($p = 0.005$). In CPRD, 16 patients were found to have a VTE complication, of which six were unique to this dataset and not found in HES (table 19).

In total 20 patients from the early group compared to 47 in the late group (3.3% vs 7.0% $p=0.005$) underwent a re-operation. In the later treatment group 70 patients (10.5%) experienced bowel obstruction which was managed conservatively compared to 52 (8.7%) in the early group. Through CPRD, 59 patients were identified as having suffered a wound infection managed conservatively, of which 57 (96.6%) were unique to this dataset.

Using HES alone, 88 (14.7%) patients from the early adjuvant chemotherapy group compared to 142 (21.3%) in the late group experienced a complication. By also including data from CPRD, the number of patients identified increased by 24% for both groups ($n=21$ for early treatment group vs $n=35$ in late treatment group).

Table 19 - Complications recorded after colonic resection and before start of adjuvant chemotherapy treatment

		Interval to adjuvant chemo		Total (%)	p value
		Before 8 weeks n (%)	8 weeks or more n (%)		
HES	Systemic ^a	20	46	66	-
	Re-operation	20 (3.3)	47 (7.0)	67 (5.3)	< 0.005
	Bowel obstruction without an operation	52 (8.7)	70 (10.5)	122 (9.6)	0.283
	Wound infection without an operation	4 (0.7)	5 (0.7)	9 (0.7)	0.895
CPRD	Systemic ^a	6	10	16	-
	Systemic not in HES ^a	2	4	6	-
	Wound infection	22 (3.7)	37 (5.5)	59 (4.7)	0.117
	Wound infection not in HES	20 (3.3)	37 (5.5)	57 (4.5)	0.060
HES and CPRD	Systemic complication	21 (3.5)	47 (7.0)	68 (5.4)	<0.005
At least one complication	HES	88 (14.7)	142 (21.3)	230 (18.2)	<0.005
	HES or CPRD	109 (18.2)	177 (26.5)	286 (22.6)	<0.0005

^aAs a patient may have experienced more than one systemic complication, percentage and Chi square p value are not provided for these categories

4.3.3.3 Regression analysis for delays

Compared with patients aged over 75, those under 60 were significantly less likely to experience a delay in adjuvant chemotherapy treatment (OR 0.65 CI 0.46 to 0.93, $p < 0.019$) while there was no difference compared to the 60 to 75 age group (table 20). Patients who smoked were more likely to experience a delay (OR 1.60 CI 1.08 to 2.35 $p = 0.018$). Gender, the year of resection, laparoscopy or the type of admission did not significantly alter the odds of a delay in adjuvant chemotherapy. If the patient received a stoma at the index operation (OR 1.79 CI 1.26 to 2.52 $p = 0.001$), or experienced a complication following resection (OR 1.53 CI 1.16 to 2.03 $p = 0.003$), the odds of starting adjuvant treatment after eight weeks or more increased.

Table 20 - Odds ratio for first adjuvant chemotherapy treatment after eight weeks or more

		Odds ratio for treatment 8 weeks or more (95% CI)	p value
Age (years)	Over 75	ref	
	60 to 75	0.73 (0.53 to 1.01)	0.056
	Under 60	0.65 (0.46 to 0.93)	0.019
Gender	Male	ref	
	Female	1.17 (0.93 to 1.46)	0.182
Smoker	No	ref	
	Yes	1.60 (1.08 to 2.35)	0.018
Charlson Score	Zero	ref	
	One or more	1.00 (0.76 to 1.32)	0.997
Surgical Access	Open	ref	
	Laparoscopic	1.12 (0.80 to 1.56)	0.514
Admission	Elective	ref	
	Emergency	1.10 (0.84 to 1.45)	0.485
Stoma	No	ref	
	Yes	1.79 (1.26 to 2.52)	0.001
Resection year	2000 to 2003	ref	
	2004 to 2007	1.04 (0.77 to 1.42)	0.786
	2008 to 2011	1.15 (0.85 to 1.57)	0.373
Complication	No	ref	
	Yes	1.53 (1.16 to 2.03)	0.003

Impact on survival

By the end of ONS coverage, 338 (26.7%) patients had died, with a significantly higher proportion in the late treatment group (n=220, 32.9% vs n=163, 27.3% respectively p = 0.028). Compared to patients who received adjuvant chemotherapy within 8 weeks, those in the delayed group were associated with a 44% (CI 1.16 to 1.79, p=0.001) increase in the hazard ratio of death (table 21). Compared to those aged over 75 years, a younger age (Under 60: HR 0.47 95% CI 0.35 to 0.63; 60 to 75 years: HR 0.59 CI 0.46 to 0.76) was associated with a reduced hazard ratio of death, while being a smoker (HR 2.26 CI 1.41 to 3.64) increased the hazard ratio.

An emergency resection significantly increased the hazard ratio by 59% (95% CI 1.28 to 1.98, p<0.0005) compared to those who underwent an elective resection. However, undergoing a laparoscopic rather than an open resection (HR 0.84 CI 0.56 to 1.27, p=0.414) or experiencing a complication (HR 1.15 CI 0.89 to 1.48, p=0.295) did not significantly alter the hazard of mortality. Those who underwent a resection between 2008 to 2011 inclusive were

associated with a significant decrease in the hazard ratio of death compared to patients treated between 2000 to 2003 (HR 0.64 CI 0.46 to 0.88 p=0.007).

Table 21 - Hazard ratio of death for patients receiving adjuvant chemotherapy

		Hazards ratio of death (95% CI)	p value
Adjuvant chemo	Within 8 weeks	ref	
	8 weeks or later	1.44 (1.16 to 1.79)	0.001
Age (years)	Over 75	ref	
	60 to 75	0.59 (0.46 to 0.76)	< 0.0005
	Under 60	0.47 (0.35 to 0.63)	< 0.0005
Gender	Male	ref	
	Female	0.82 (0.67 to 1.01)	0.063
Smoker	No	ref	
	Yes	2.26 (1.41 to 3.64)	0.001
Charlson Score	Zero	ref	
	One or more	1.12 (0.88 to 1.43)	0.361
Surgical Access	Open	ref	
	Laparoscopic	0.84 (0.56 to 1.27)	0.414
Admission	Elective	ref	
	Emergency	1.59 (1.28 to 1.98)	< 0.0005
Resection year	2000 to 2003	ref	
	2004 to 2007	0.89 (0.70 to 1.34)	0.351
	2008 to 2011	0.64 (0.46 to 0.88)	0.007
Complication	No	ref	
	Yes	1.15 (0.89 to 1.48)	0.295

Figure 23 (and life table 22) suggests a delay in adjuvant chemotherapy was associated with a relative increase in survival by 12.3% at 5 years if the patients had received treatment within eight weeks (65% in the early treatment group vs 73% in the late treatment group).

Figure 23 - Kaplan Meier curves for first adjuvant chemotherapy episode

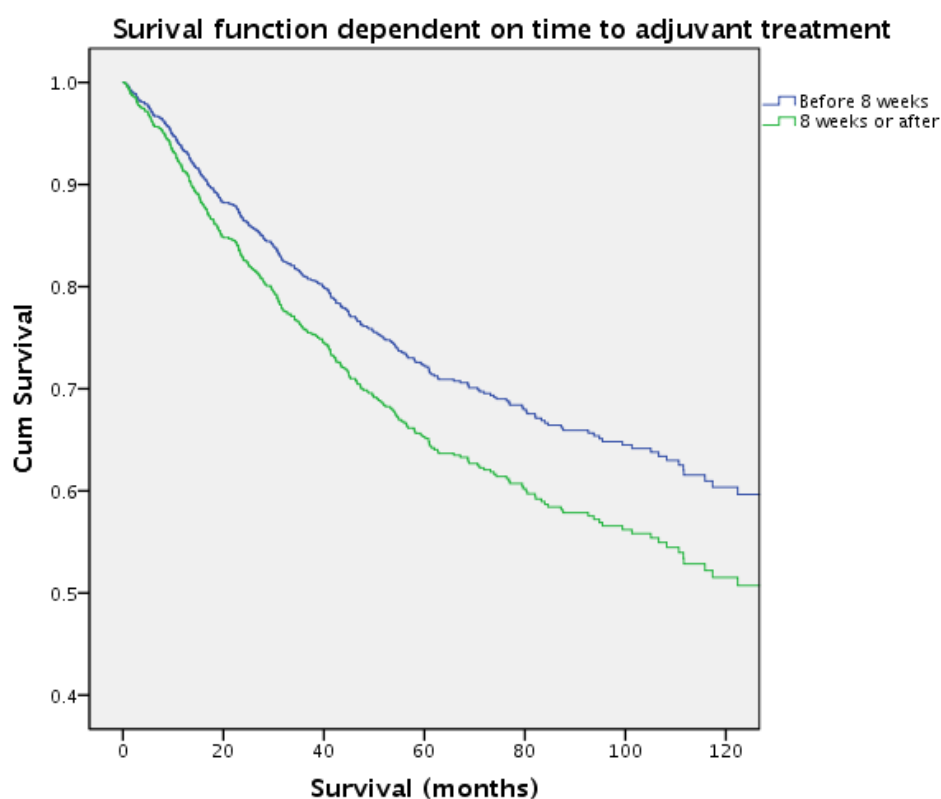


Table 22 - Life table stratified according to first adjuvant chemotherapy episode

First episode of AC	Survival (months)										
	0	12	24	36	48	60	72	84	96	108	120
Before 8 weeks	598	481	395	314	242	185	148	114	93	68	47
After 8 weeks or more	668	516	397	295	209	152	115	87	67	48	28
Total	1266	997	792	609	451	337	263	201	160	116	75

4.3.4 Discussion

4.3.4.1 Summary

The aim of this study was to determine the factors associated with a delay in initiating adjuvant chemotherapy following colonic cancer resection, and their impact on long term prognosis. The analysis suggests a smoking history, stoma at the time of resection and older

age are associated with a delay in adjuvant chemotherapy after eight weeks. After adjusting for age, emergency admission and a smoking history, a delay in adjuvant chemotherapy after eight weeks is associated with a poorer prognosis. Open surgery or an emergency admission were not associated with delay in treatment, but the latter independently reduced overall survival. CPRD was shown to include a high proportion of wound infections and VTE events that were not noted in HES. The findings of this study suggest complications can lead to a poorer prognosis by delaying the start of adjuvant chemotherapy, rather than directly affecting survival.

4.3.4.2 Comparison to previous studies

This is the first study to use routinely collected national primary and secondary care data to identify a combination of factors that are associated with a delay in adjuvant chemotherapy. One of the co-authors of this study demonstrated CPRD can also provide a source of post-operative complications that may have been solely managed in the community, such as venothromboembolic events¹⁴⁹.

In this study, the overall survival for colon cancer patients treated within 8 weeks was found to be 65% at 5 years with a HR of 1.44 (1.16 to 1.79). Nachiappan et al.¹⁴⁸ demonstrated a 5 year OS of about 68% (extrapolated from graph) using HES data and similar definitions as stated in this study. Ahmed et al.¹⁵⁰ described a 5 year overall survival of 59.5% using medical records for colorectal cancer patients identified from the cancer registry, with a HR of 1.02 (95% CI of 0.80 – 1.30). Bayraktar et al.¹⁵¹ used hospital records from a single centre to demonstrate a 5 year OS of 78.6% for patients treated within 60 days and a HR for OS of 2.07 (95% CI 1.02 – 4.19). Cheung et al.¹⁵² sourced patients from the SEER database and found treatment within 2 months was associated with a 5 year OS of 61.5%, while Wadd et al.¹⁵³ showed a 5 year OS of 65.8% using hospital records.

The number of colon cancer patients' cases identified through my method compares well to the National Bowel Cancer Audit¹⁵⁴. In 2011 the latter reported 17161 patients nationally underwent a colonic resection. From our data in 2010, 1242 had a resection, representing 7.2%. This is close to the 8% coverage of the national population by CPRD. However the proportion of patients receiving adjuvant treatment could be considered low, as previous

studies have shown almost 20% of colon cancer patients receive treatment¹⁴³. This is likely related to the completeness of chemotherapy coding in HES which has improved in latter years.

The overall re-operation rate is lower than the 14% reported by van der Geest et al.¹³⁴ but similar to the 5.6% reported by Nachiappan et al. This reflects the different data sources: Geest et al. used medical records, whereas the latter group used Hospital Episodes Statistics similar to our study.

These comparisons demonstrate the linked HES data provided by CPRD is similar to standard HES and patients records. It also suggests patients extracted from CPRD reflect the wider cancer population.

Although it is an expected finding that complications are associated with a delay in the first episode of adjuvant chemotherapy, I was able to describe in detail complications from both primary and secondary care. For example, in previous studies post-operative hospital stay has been used as a surrogate marker for complications managed conservatively¹⁵⁵, but data for conservative treatment of bowel obstruction has not previously been published. In my study this was the most common complication (9.6% n=122), and in the majority is likely to represent post-operative ileus. By utilising CPRD, we identified 56 extra patients as having suffered a complication, with the majority being wound infections. 5.3% (n=67) of all patients underwent a re-operation, with a significantly higher proportion in the delayed adjuvant chemotherapy group (7.0%, n=47).

In addition to complications, I found a stoma at the initial operation was a risk factor for delays. However it was difficult to ascertain whether the stoma itself, rather the type of operation led to the delay. Interestingly smoking was found to be an independent risk factor for a delay in adjuvant treatment. An interaction between smoking and complications or Charlson score was not demonstrated to explain this finding. A surgeon may decide to use a minimally invasive approach to reduce recovery time¹⁵⁶. However, this study demonstrates there is no difference in using laparoscopy or open surgery in terms of reducing post-operative treatment delays.

My finding that complications do not independently reduce survival is similar to a previous study by Tevis et al.¹⁴⁷ which included data from rectal cancer patients. Nachiappan et al.¹⁵⁷ also found that it was the delay in treatment, rather than the presence of a re-operation that led to worse prognosis. As expected a smoking history¹⁵⁸ or an emergency admission⁹⁰ reduced overall survival. Patients treated between 2008 to 2011 demonstrated a better prognosis than those treated in 2000-2003. This could reflect a change in practice to treat high risk stage II colon cancer with adjuvant chemotherapy, or the addition of oxaliplatin in chemotherapy regimens¹⁵⁹.

4.3.4.3 Strengths and Limitations

This is the first study to use linked data from CPRD as the source and so I have been able to consider post-discharge complications leading to delays in treatment. I am unable to ascertain the percentage of patients that had a delay in treatment due to administrative issues unrelated to the resection, such as postponed clinics or unavailability of services. Also, I could not determine whether a patient completed the treatment regimen. The latter has been shown to effect overall survival¹⁶⁰. As I did not include stage, I assumed the presence of adjuvant chemotherapy episodes, and the exclusion of patients with a code indicating metastasis was sufficient to identify patients receiving adjuvant chemotherapy for curative purposes.

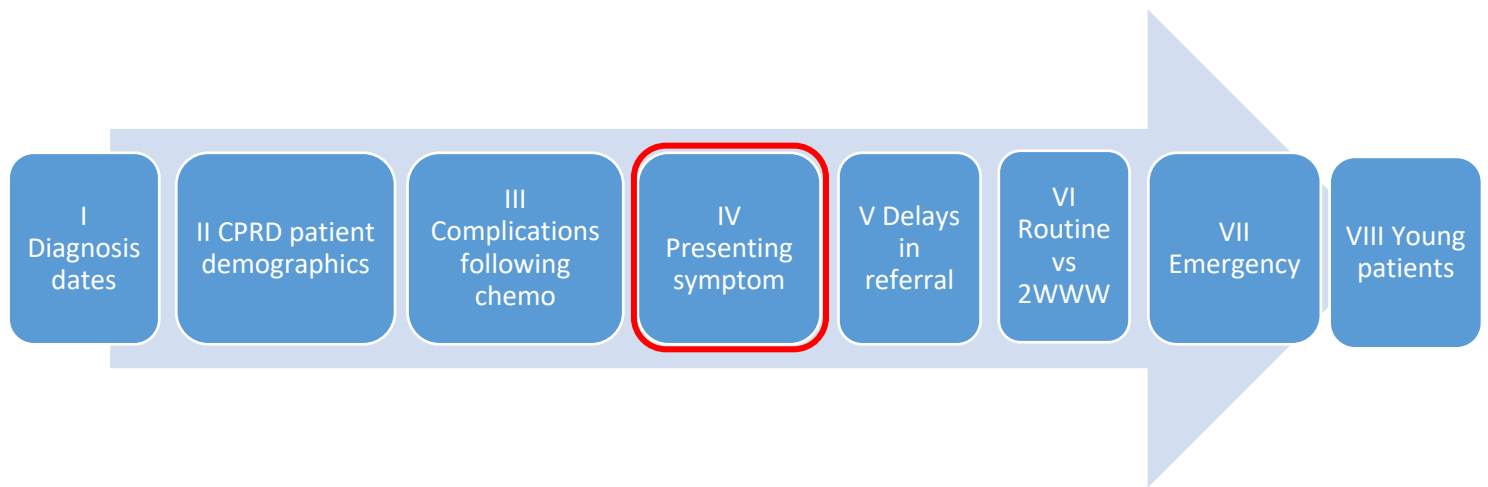
4.3.4.4 Conclusion

After adjusting for potential confounders such as type of surgical access and emergency admission, I have demonstrated complications, identified in primary care and in the hospital, are associated with a delay in adjuvant chemotherapy following resection for colon cancer. The most common complication identified from the GP data was wound infections. However, complications did not independently reduce prognosis. The significant reduction in survival following a delay in adjuvant chemotherapy should encourage the clinician to ensure all steps are taken to reduce complications and manage them effectively.

4.3.4.5 Conclusion related to thesis

1. Cancer patients extracted using CPRD data adequately represent the general cancer population
2. Timing of key clinical events in routinely collected national data reflects findings using other sources
3. CPRD linked data can be used to define delays and outcome

End of Study



4.4 Study IV - Presenting symptoms for non-emergency colorectal cancer patients

In the previous chapters I have determined CPRD is a reliable dataset to determine delays along the cancer patient pathway, and is representative of the national population. In this chapter I describe the most common symptoms recorded at the first presentation for cancer patients diagnosed as a non-emergency.

The hypotheses for this study are:

- **The type of symptom at presentation for non-emergency patients varies according to gender, age and co-morbidities**
- **Combination of symptoms are poorly recorded in CPRD**

I aim to answer the following questions:

- **Are certain patient demographics associated with particular symptoms?**
- **What is the extent of missing data in regards to symptoms within CPRD?**

4.4.1 Introduction

Although previous studies have described the symptoms associated with certain patient demographics, none have been able to consider possible confounders that may influence the likelihood of a particular symptom. By understanding the type of symptom at presentation for cancer patients, it may improve the detection of patients at an earlier stage.

As described previously, consultation data provided by CPRD contains only READ codes, without free text. It is possible that the combination of symptoms recording in CPRD is poor as symptom deemed less severe or relevant may have been written in the free text. By understanding such limitations of CPRD, it is possible to adapt the methods to provide reliable results regarding delays.

4.4.2 Method

4.4.2.1 Patient inclusion

All the patients included had a diagnosis confirmed in the cancer registry between 2000 – 2013. Patients diagnosed via an emergency, death certificate only and screening were excluded as it is unclear the level of GP involvement in their diagnosis. Those with a previous cancer diagnosis recorded in any one of the three datasets were also excluded as the GP would have been more aware of a potential cancer diagnosis in these patients. The demographics for these patients, including the number of patients per practice was described.

4.4.2.2 Definition of symptoms before diagnosis

Relevant symptoms included have been previously described in chapter 3.2. The presenting symptom was defined as the first symptom recorded in CPRD in the year before colorectal cancer diagnosis as noted in the Cancer Registry.

4.4.2.3 Comparison of patients with and without a symptom recorded before diagnosis

Patients with a symptom recorded in CPRD in the year before the cancer diagnosis were compared with those without a diagnosis using univariate and multivariate analysis. Using a binary logistic regression model, the latter described the odds ratio of a missing symptom

for each category with female, age 60 – 69, IMD 1, Charlson score zero, year 2006 – 08 and stage IV the relevant reference category.

4.4.2.4 Presenting symptoms in the year before diagnosis

The percentage of patients presenting with a particular symptom alone or in combination was described. The median and interquartile range in days from first presentation to diagnosis was calculated for each symptom, with significance differences between the absence or presence of that symptom tested with the Mann U Whitney test ($p < 0.05$).

A binary logistic regression analysis was carried out to calculate the odds ratio of the presence of a symptom at first presentation according to gender, age, Charlson score, IMD and tumour site. This model was repeated for each symptom individually.

4.4.3 Results

4.4.3.1 Patient cohort

Of the **22019** colorectal cancer patients with a diagnosis in the cancer registry between 2000 to 2013 inclusive, 5232 (23.8%) were diagnosed through an emergency

↳ **16787** non-emergency patients, 261 were diagnosed through DCO, identified from the cancer registry or the route to diagnosis algorithm.

↳ **16526** patients, 901 patients were excluded due to the screening route. As screening was not a national policy before 2006, a small number, if any, would have been diagnosed through this route before 2006.

↳ **15625** patients, 1522 (9.7%) patients were diagnosed with another cancer in one of the three datasets preceding the CRC diagnosis in the CR.

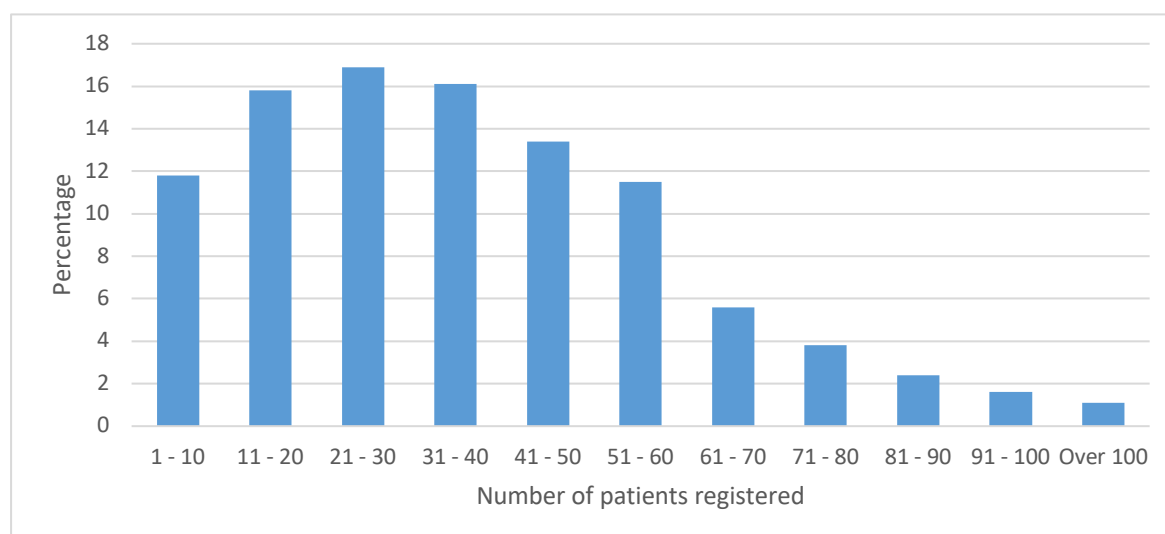
For the remaining **14103** patients, the demographics are below (table 23).

Table 23 - Demographics and tumour characteristics for 14103 patients with a recording of colorectal cancer in the Cancer Registry between 2000 and 2013

		N	%
Gender	Male	7826	55.5
	Female	6277	44.5
Age	Under 50	667	4.7
	50 – 59	1755	12.4
	60 – 69	3251	23.1
	70 - 79	4724	33.5
	80 and over	3706	26.3
Charlson	Zero	11105	78.7
	One	1873	13.3
	Two or more	1125	8.0
IMD	1	3477	24.7
	2	3547	25.2
	3	2913	20.7
	4	2447	17.4
	5	1719	12.2
Stage	I	1742	12.4
	II	3689	26.2
	III	3583	25.4
	IV	1783	12.6
	Missing	3306	23.4
Site	Colon	9427	66.8
	Rectum	4676	33.2

These patients were registered with 373 practices, ranging from two patients to 161 patients per practice, with an average of 36.9, median of 33 (IQR 18 – 51) and a mode of 26 patients. Figure 24 describes the percentage of patients per practice.

Figure 24 - Percentage of patients registered with each of the 373 practices



4.4.3.2 Comparison of patients with and without a symptom recorded in the year before diagnosis

10661 (71.6%) patients had a relevant symptom recorded at least once in the year before the diagnosis in the Cancer Registry.

The first comparison is between patients with and without a symptom recorded in the year before diagnosis (table 24). There was no significant difference in gender or age. However patients diagnosed in the latter years were more likely to have symptoms recorded in the year leading up to diagnosis: in 2000 – 01 28.9% did not have symptom recorded as opposed to 23.6 in 2012 – 13. Patients with a Charlson score of more than two, or those from a lower socioeconomic area, were also less likely to have a symptom recorded before diagnosis. In multivariate analysis, male patients (OR 0.91 95% CI 0.85 – 0.99) were less likely to have a symptom recorded in CPRD in the year leading up to diagnosis, as were patients with a Charlson score of two or more compared to those with a score of zero (OR 0.82 95% CI 0.71 – 0.95 $p = 0.007$). Patients diagnosed with stage II or III were more likely to have a symptom recorded compared to those with stage IV, but no difference was seen with stage I patients. On the contrary missing stage patients were less likely to have a symptom recorded. Compared to 2006 – 08, only diagnosis in 2000-02 was less likely to be associated with a symptom (OR 0.72 95% CI 0.61 – 0.85 $p < 0.0005$).

Table 24 – Comparison of patient demographics between patients with and without a symptom recorded in the year leading up to cancer diagnosis – univariate and multivariate analysis

		Symptom present in the year before cancer diagnosis				Odds ratio of symptom recorded year before cancer diagnosis		
		Yes		No		p	OR (95% CI)	p
		n	%	n	%			
Gender	Male	5870	75.0	1956	25.0	0.070	0.91 (0.85 – 0.99)	0.025
	Female	4791	76.3	1486	23.7		ref	
Age	Under 50	496	74.4	171	25.6	0.197	0.91 (0.74 – 1.13)	0.382
	50 - 59	1335	76.1	420	23.9		0.98 (0.85 – 1.12)	0.765
	60 - 69	2475	76.1	776	23.9		ref	
	70 - 79	3603	76.3	1121	23.7		1.02 (0.92 – 1.14)	0.678
	80 and over	2752	74.3	954	25.7		0.97 (0.85 – 1.10)	0.648
Charlson score	Zero	8433	75.9	2672	24.1	0.024	ref	
	One	1415	75.5	458	24.5		0.95 (0.85 – 1.07)	0.403
	Two or more	813	72.3	312	27.7		0.82 (0.71 – 0.95)	0.007
IMD	1	2714	78.1	763	21.9	< 0.0005	ref	
	2	2681	75.6	866	24.4		0.86 (0.75 – 0.99)	0.031
	3	2163	74.3	750	25.7		0.81 (0.69 – 0.95)	0.008
	4	1858	75.9	589	24.1		0.90 (0.79 – 1.06)	0.197
	5	1245	72.4	474	27.6		0.77 (0.65 – 0.92)	0.003
Stage	I	1303	74.8	439	25.2	< 0.0005	0.99 (0.83 – 1.19)	0.944
	II	2925	79.3	764	20.7		1.31 (1.13 – 1.51)	< 0.0005
	III	2798	78.1	785	21.9		1.20 (1.04 – 1.39)	0.012
	IV	1332	74.7	451	25.3		ref	
	Not provided	2303	69.7	1003	30.3		0.81 (0.69 - 0.96)	0.014
Year of diagnosis	2000 – 02	1729	71.1	704	28.9	< 0.0005	0.72 (0.61 – 0.85)	< 0.0005
	2003 – 05	2382	74.7	805	25.3		0.87 (0.77 – 1.01)	0.052
	2006 – 08	2604	77.2	771	22.8		ref	
	2009 – 11	2518	77.7	721	22.3		1.00 (0.89 – 1.12)	0.955
	2012 – 13	1428	76.4	441	23.6		0.91 (0.79 – 1.06)	0.238

The regression model includes all other variables described here

4.4.3.3 Symptom at first presentation in the year leading up to diagnosis

The following table (table 25) describes the symptoms recorded at the first presentation in the year leading up to a colorectal cancer diagnosis. The most common symptom was rectal bleeding (19.5%) followed by abdominal pain (14.6%) and anaemia (11.5%). It is apparent that in the majority of cases, there is only one symptom recorded, especially for rectal bleeding (97.2%) and anaemia (97.2%). Abdominal mass and weight loss were most commonly noted in combination with another symptom, namely change in bowel habit (6.1%) or abdominal pain (6.1%) for the former and abdominal pain (4.7%) for the latter.

Table 25 - Percentage of patients presenting with each symptom at the first presentation before a colorectal cancer diagnosis

	Alone		Rectal bleeding		Anaemia		CBH		Diarrhoea		Abdominal pain		Mass		Constipation		OBH		Fatigue		Weight Loss			
	N	% of total	n	% of N	n	% of N	n	% of N	n	% of N	n	% of N	n	% of N	n	% of N	n	% of N	n	% of N	n	% of N		
Rectal bleeding	2756	19.5	2678	97.2			3	0.1	22	0.8	11	0.4	20	0.7	8	0.3	11	0.4	3	0.1	6	0.2	8	0.3
Anaemia	1616	11.5	1572	97.2	3	0.2			5	0.3	4	0.2	8	0.5	3	0.2	6	0.4	0	0.0	10	0.6	10	0.6
CBH	1235	8.8	1166	94.4	22	1.8	5	0.4			9	0.7	12	1.0	9	0.7	3	0.2	2	0.2	5	0.4	7	0.6
Diarrhoea	1231	8.7	1185	96.3	11	0.9	4	0.3	9	0.7			14	1.1	3	0.2	1	0.1	1	0.1	5	0.4	6	0.5
Abdominal pain	2052	14.6	1953	95.2	20	1.0	8	0.4	12	0.6	14	0.7			9	0.4	26	1.3	7	0.3	5	0.2	15	0.7
Abdominal mass	147	1.0	112	77.8	8	5.4	3	2.0	9	6.1	3	2.0	9	6.1			6	4.1	1	0.7	0	0.0	5	3.4
Constipation	866	6.1	811	93.6	11	1.3	6	0.7	3	0.3	1	0.1	26	3.0	6	0.7			4	0.5	4	0.5	5	0.6
OBH	242	1.7	227	93.8	3	1.2	0	0.0	2	0.8	1	0.4	7	2.9	1	0.4	4	1.7			1	0.4	1	0.4
Fatigue	472	3.3	433	91.7	6	1.3	10	3.1	5	1.1	5	1.1	5	1.1	0	0.0	4	1.6	1	0.3			5	1.1
Weight loss	321	2.3	273	85.0	8	2.5	10	2.1	7	2.2	6	1.9	15	4.7	5	1.6	5		1	0.2	5	1.6		

CBH = Change in Bowel Habit, OBH = Other Bowel Habit

4.4.3.4 Odds ratio for presentation with each symptom

Male patients were more likely to present with a change in bowel habit and diarrhoea, but less likely to present with anaemia, abdominal mass, abdominal pain, other bowel function and fatigue (table 26 and 27). The youngest age group (under 50 years of age) are more likely to present with abdominal pain and 'other bowel function' than those aged 60 – 69. With increasing age, anaemia, diarrhoea, constipation or fatigue becomes more common as presenting symptoms. There is no significant difference with the other age groups. Anaemia, weight loss and fatigue are associated with a colonic tumour as opposed to a rectal site. This is in contrast to rectal bleeding, change in bowel habit and diarrhoea, all of which are more likely to be a presenting symptom for rectal tumours.

Red-flag symptoms at presentation, such as rectal bleeding, change in bowel habit and the presence of a mass, significantly reduced the diagnostic interval, whereas anaemia and diarrhoea increased this interval. All non-specific symptoms were associated with a longer diagnostic interval, except weight loss which showed no significant difference (table 26 and 27).

Table 26 The odds ratio of presenting with a particular red-flag symptom, according to age, gender and tumour site and the diagnostic interval for that particular symptom

		Rectal bleeding	Anaemia	Change bowel habit	Diarrhoea	Mass
Gender	Male	1.06 (0.96 – 1.16)	0.89 (0.80 – 0.99)***	1.53 (1.34 – 1.76)***	1.31 (1.17 – 1.47)**	0.64 (0.44 – 0.91)**
	Female	ref				
Age	Under 50	1.02 (0.81 – 1.28)	0.65 (0.45 – 0.95)*	0.52 (0.35 – 0.76)***	0.91 (0.67 – 1.25)	1.51 (0.67 – 3.39)
	50 – 59	1.18 (1.01 – 1.37)*	0.72 (0.56 – 0.93)*	0.79 (0.64 – 0.97)	0.84 (0.67 – 1.06)	1.27 (0.66 – 2.45)
	60 – 69	ref				
	70 - 79	0.77 (0.68 – 0.88)***	1.46 (1.24 – 1.72)***	0.94 (0.80 – 1.11)	1.12 (0.96 – 1.30)	1.18 (0.71 – 1.95)
	80 and over	0.71 (0.62 – 0.82)***	2.05 (1.74 – 2.42)***	0.74 (0.62 – 0.88)*	1.24 (1.06 – 1.44)**	1.42 (0.87 – 2.31)
Site	Colon	0.35 (0.32 – 0.39)***	4.64 (3.90 – 5.51)***	0.51 (0.45 – 0.58)***	0.60 (0.53 – 0.68)***	1.20 (0.81 – 1.77)
	Rectum	ref				
Time to diagnosis	Presence	50 (21 – 108)***	81 (37 – 180)***	43 (21 – 86)***	83 (40 – 172)***	35 (16 – 67)***
	Absence	82 (37 – 179.5)	71 (31 – 155)	78 (34 – 170.25)	71 (31 – 157)	73 (32 – 160)

The diagnostic interval was compared for patients with and without the symptom recorded in CPRD, significant difference tested with the Mann-U Whitney test. *** < 0.0005 **0.0005 – 0.005 *0.005 – 0.05

Table 27 – The odds ratio of presenting with a particular non-specific symptom, according to age, gender and tumour site and the diagnostic interval for that particular symptom

		Abdominal pain	Constipation	Other bowel	Weight loss	Fatigue
Gender	Male	0.82 (0.74 – 0.90)**	0.92 (0.79 – 1.07)	0.69 (0.53 – 0.91)***	1.06 (0.84 – 1.33)	0.65 (0.52 – 0.80)***
	Female	ref				
Age	Under 50	1.25 (1.02 – 1.60)*	1.24 (0.83 – 1.85)	2.63 (1.65 – 4.20)***	1.05 (0.50 – 2.20)	1.15 (0.67 – 1.99)
	50 – 59	1.05 (0.88 – 1.25)	1.02 (0.76 – 1.38)	1.73 (1.18 – 2.53)***	1.08 (0.71 – 1.66)	1.34 (0.94 – 1.90)
	60 – 69	ref				
	70 - 79	0.73 (0.64 – 0.84)***	1.68 (1.38 – 2.05)***	0.70 (0.49 – 1.02)	1.54 (1.13 – 2.11)**	1.29 (0.97 – 1.71)
	≥ 80	0.48 (0.41 – 0.56)***	2.26 (1.79 – 2.85)***	0.56 (0.37 – 0.86)	1.33 (0.94 – 1.88)	1.55 (1.19 – 2.02)***
Site	Colon	3.41 (3.00 – 3.88)***	0.91 (0.77 – 1.06)	0.92 (0.70 – 1.21)	1.73 (1.33 – 2.26)***	1.85 (1.48 – 2.32)***
	Rectum	ref				
Time to diagnosis	Presence	94 (6 – 197)***	112.5 (49 – 232)***	152.5 (81.5 – 239.5)***	66 (33.5 – 170.5)	121 (59 – 239)***
	Absence	68 (29 – 149)	70 (31 – 152)	71 (31 – 155)	73 (32 – 159)	70 (31 – 155)

*The diagnostic interval was compared for patients with and without the symptom recorded in CPRD, significant difference tested with the Mann-U Whitney test. *** < 0.0005 **0.0005 – 0.005 *0.005 – 0.05*

4.4.4 Discussion

4.4.4.1 Comparison to other studies

Overall nearly a third of patients did not have a symptom recorded. Patients with a diagnosis in the earlier years within CPRD are more likely to have a missing symptom in the year preceding diagnosis. There are two possible explanations for this. This could be a reflection of the quality of symptom recording within CPRD which may have improved in more recent years. Alternatively this could reflect the introduction of referral guidelines of 2005. In a study by Price et al.¹²⁴ the authors compared the free text from consultations to the medcodes and found information regarding non-specific symptoms were more likely to have missing medcodes. The symptoms regarded as red-flag in this study are based on the 2005 NICE criteria. Therefore the higher level of missing data before 2006 may be explained by what the GP felt was relevant to cancer. The year of diagnosis should be taken into account in regression modelling if data from before 2006 is used.

4.4.4.2 Relevance to clinical practice

This is the largest study to describe the patient demographics associated with the presence of a particular symptom. It is apparent that young patients are more likely to present with non-specific symptoms, a combination of which may contribute to their later stage at diagnosis and higher rate of emergency presentations (this is further explored later). As with any study into symptomology it is not possible to determine whether the symptom is

actually due to the underlying cancer. Anaemia was found to be more common in the elderly and female patients, which is likely a reflection of an underlying non-cancer pathology rather than the cancer itself. However the remaining symptoms in the year leading up to diagnosis is likely due to the underlying cancer. More importantly this describes the consultations encounters with a GP, who then has the difficult task of differentiating between cancer and non-cancer patients based on these symptoms.

Interestingly patients with stage I disease were equally likely to have a symptom recorded in the year before diagnosis as patients with stage IV disease. As all these patients were deemed to have been diagnosed via the GP by excluding DCO and screening, they all should have a symptom recorded. The study above that compared medcodes to free text could also explain the higher level of missing data for stage IV – there are more likely associated with non-specific symptoms. However the lack of symptoms recording for Stage I disease is not explained by this.

4.4.4.3 Limitations of dataset

There is a paucity of data regarding the combination of symptoms, if only the first presentation is considered. This is likely to reflect the method of recording symptoms by the GP during the consultations. The symptoms most associated with cancer, such as rectal bleeding and change in bowel habit are likely to be recorded as a medcode, with secondary non-specific symptoms such as abdominal pain being recorded as free text, the latter which is not available to researchers. Consequently consultations with only non-specific symptoms in CPRD are unlikely to have red-flag symptoms. On the contrary, consultation with red-flag symptoms recorded may have had non-specific symptoms which have been lost in the free text.

GP practices voluntarily provide their data to CPRD, with a certain proportion of these allow their patients' data to be linked to the cancer registry. Individual patients are able to withdraw consent for their data being available for research, although this is thought to be a minority. Therefore practices with lower number of cases are smaller practices in terms of patients registered rather than a reflection of missing data. The use of the Generalised Estimating Equation module in SPSS attempts to overcome the potential bias created by

practices with a high number of patients.

4.4.4.4 Relevance to thesis

The findings above provide preliminary results regarding the completeness of symptom recording in CPRD and the patient characteristics associated with certain symptoms. This analysis provides some credence to CPRD and my methodology, while highlighting potential limitations that need to be taken into account. For the lack of combination of symptoms the options are to:

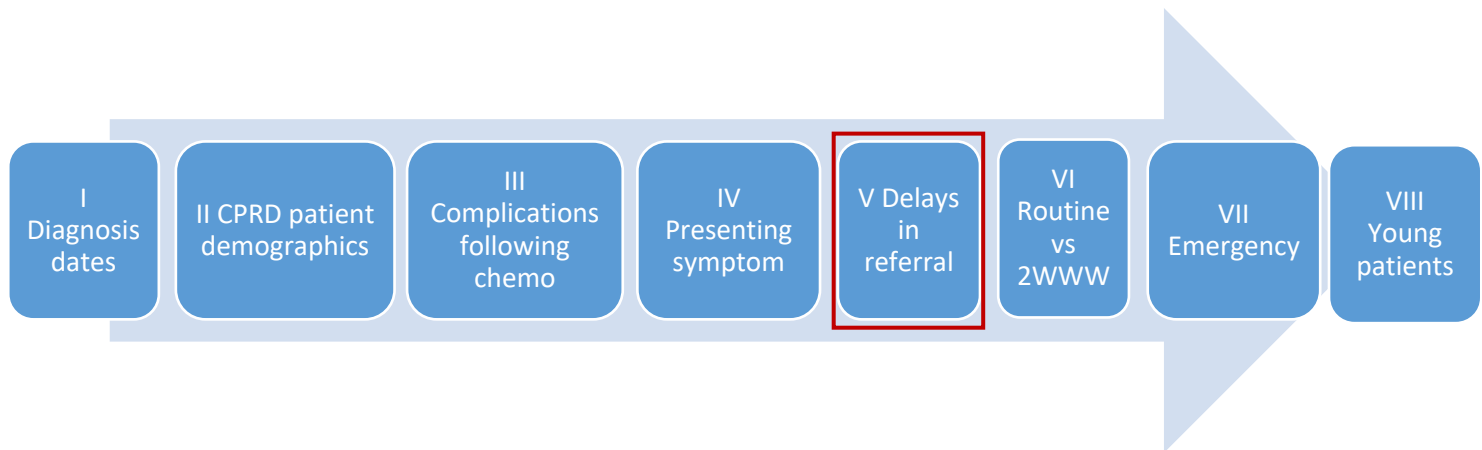
1. Group patients as presenting with individual or combination symptoms. Therefore patients with rectal bleeding symptoms will be classified as either rectal bleeding, bleeding with anaemia, bleeding with change in bowel habit etc. However there are too few patients with combination symptoms to carry out meaningful comparisons.
2. Group patients according to the positive predictive value of a colorectal cancer diagnosis. The NICE referral guidance is based on this, and therefore should determine the urgency of referral by the GP. As such the order would be:

Rectal bleeding, abdominal mass, anaemia, change in bowel habit, diarrhoea, abdominal pain, constipation, other bowel function, weight loss and fatigue

The GP is likely to have recorded the symptom of most concern, in terms of the likelihood it is due to cancer. This would explain the higher percentage of patients with rectal bleeding alone (97.2%), as compared with those with non-specific symptoms (for example other bowel function 93.8%, fatigue 91.7% and weight loss 81.0%). Therefore I have chosen the second classification above.

The following chapters are designed to further our understanding of the causes and impact of delays in primary care.

End of Study



4.5 Study V – Are delays in primary care associated with a late stage and worse outcome?

In the previous chapter I described the symptoms associated with a non-emergency colorectal cancer diagnosis, and how this differs according to patient demographics. In the following chapter I determine how the presenting symptom impacts delays in referral and diagnosis, and whether these delays worsen prognosis.

The hypothesis of this study states:

- **Delays in referral from primary care are associated with a late stage of colorectal cancer and a worse prognosis.**

The aims of this study are to answer the following questions:

- **Which presenting symptoms are associated with delays in referral from primary care?**
- **How does the hospital interval differ according to the presenting symptom?**
- **Does the presenting symptom alter the stage at diagnosis?**
- **Are delays in primary care associated with a later stage of cancer?**
- **Are delays in primary care associated with a worse prognosis?**

4.5.1 Introduction

This study investigates the central aim of my thesis. Delays in referral are considered using two metrics – time in days and number of consultations. The hypothesis and the aims are the same for both. Data from the national audit of primary care suggests there is an overlap in the IQR of the time elapsed for referrals after two (IQR 7 – 40) or three (IQR 15 – 75.5) consultations. Tumour progression will continue with time, irrespective of the number of consultations before referral, suggesting delays should be based on time rather than consultations. However the number of consultations provides a tangible target for health policy.

4.5.2 Method

4.5.2.1 Patient cohort

Patients included had been diagnosed with colorectal cancer between 2000 – 2013 inclusive in the cancer registry and had a referral based on the READ codes (see appendix vi) in the year leading up to diagnosis. Colon and rectal cancers were analysed separately.

4.5.2.2 Odds of referral data recorded

All patients with a referral were compared to those with a referral to understand whether there is a bias in CPRD for certain patient groups not have a referral recorded. A binary logistic regression was created, with age, gender, IMD, year of diagnosis, Charlson score and stage included as independent variables. A sensitivity analysis by only including patients diagnosed from 2006 onwards is carried out to ensure the year of diagnosis does not impact the inclusion of referral data.

4.5.2.3 Symptoms of interest

Symptoms relevant to colorectal cancer are described previously in chapter 3.2 The first presentation was considered the first consultation with a relevant symptom in the year before the referral. The referral was considered the end-point of the year to determine the relevant symptoms in this study, rather than the date of diagnosis, as the GP would have made a referral based on the presentations leading up to this decision. Patients were represented by each symptom as per the previous chapter.

4.5.2.4 Relevant intervals along patient pathway

The median and interquartile range for the primary care interval (first presentation to referral), hospital interval (referral to diagnosis) and diagnostic interval (first presentation to diagnosis) were calculated for each symptom, and for red-flag and non-specific group of symptoms. Significant differences in the intervals were compared to rectal bleeding using the Mann Whitney U test.

4.5.2.5 Stage at diagnosis

The stage at diagnosis (I, II, III, IV or missing) was described according to the presenting symptom. Significant differences in the proportion of patients with each stage were tested for by comparing the residuals at the 0.05 significance level.

4.5.2.6 Delays in referral analysis with delays measured as time

In a national audit up to 90% of patients with colorectal cancer were referred within three months¹⁶¹, while an interval of up to two weeks is similar to the median time for two consultations before referral³⁹. Therefore the primary care interval was split into three groups, within two weeks (Group I), two weeks to three months (Group II) and over three months (Group III), to determine whether a delay in referral was associated with a later stage at diagnosis. This was repeated with each symptom considered individually. The hospital interval was compared between the three groups above to determine if a delay in primary care was associated with a subsequent delay in hospital. This was carried out for each symptom individually.

A multinomial regression model was employed, with the referral interval as the dependent variable, to determine the patient demographics associated with a delay in referral. Referral within two weeks was termed the reference category, with age, IMD, Charlson score, year of diagnosis and gender entered as independent categorical variables.

Cox proportional hazard modelling was used to determine the hazards ratio of colorectal cancer related death following a referral between two weeks to three months and over three months compared to patients who were referred within two weeks, while accounting for age, gender, IMD and Charlson score. This was repeated by only including patients

presenting with the same symptom, to understand whether delays rather than the presenting symptom can independently contribute to a worse prognosis.

4.5.4.7 Prognosis according to presenting symptom

Cox proportional hazard modelling was used to describe the risk of death for a patient presenting with a symptom compared to those with rectal bleeding, after adjusting for age, Charlson score, IMD and gender. This was repeated after only including patient referred within two weeks to negate the effect of delays in referral. Patients with non-specific symptoms were grouped together and compared with those with red-flag symptoms.

4.5.4.8 Delays in referral analysis with delays measured according to consultations

The number of relevant consultations before referral were counted as follows. Referral after one consultation was defined as a referral date within two days of the first presentation. For the remaining patients the number of consultations was identified between the first presentation and the referral date. Duplicates based on consultation identification number and consultation date were excluded.

A significant difference in the number of consultations was checked for each symptom using Chi X analysis ($p < 0.05$). The median number of days and the interquartile range is described for increasing number of consultations. The Spearman rank test was used to determine the strength of the association between the number of consultations before referral and the time to referral expressed in days. Although this would be expected, this cannot be assumed. If there were a high proportion of missing consultations within CPRD following the first presentation, either because the consultation was not originally recorded by the GP or the Read codes used in this thesis fail to identify all relevant consultations, then patients referred after three consultation may incorrectly be labeled as referred after the first consultation. For such cases the referral interval would seem unusually long.

The Receiver Operating Characteristic (ROC) analysis was applied to determine the binary categorical variable which most accurately corresponds to interval to referral. For this the following models were used:

Model a: At initial or two vs the rest

Model b: At initial + two + three vs the rest

This allowed the number of consultations to be entered into two models. First as the dependent variable in a binary logistic model, to determine the patient characteristics that are associated with a delay in referral. Secondly into a cox regression model as an independent variable, while accounting for gender, IMD, age and Charlson score, to determine whether delays in referral were associated with a worse outcome. Both models included the presence of a red-flag symptom as a binary category. The Cox regression model was repeated after inclusion of patients with the same presenting symptom, to determine if delays were associated with a worse survival.

4.5.3 Results

4.5.3.1 Comparison of patients with and without a referral in CPRD

Of the original 14103 colorectal cancer patients with a diagnosis between 2000 – 2013 identified previously, 9113 (64.6%) patients had a referral recorded in CPRD in the year leading up to colorectal cancer diagnosis. Of these, 7130 (78.2%) had a relevant consultation recorded in the year leading up to referral, split further into 4527 (63.5%) patients with colon cancer and 2603 (36.5%) with rectal cancer. Binary logistic regression to determine the odds ratio of the presence of a referral reveals patients with an earlier stage cancer were more likely to have received a referral to secondary care compared with stage IV cancers. A higher Charlson score, colon cancer, and lower level of deprivation were associated with a lack of a referral in CPRD, as was a diagnosis in 2000 to 2005 (table 28). Age and gender had no impact on the recording of a referral. A sensitivity analysis was carried out by restricting diagnoses to 2006 onwards. This included only patients who had a confirmed elective route of diagnosis, as determined by the Elliss et al.²⁹ algorithm described above. This confirmed the findings for the larger cohort, except the odds ratio for an IMD of two, three or five were no longer significant.

Table 28 - Odd ratio of a referral according to patient demographics and tumour characteristics. All variables included in the model. Sensitivity analysis for diagnoses from 2006 onwards included

		Year of diagnosis	
		2000 - 2013	2006 - 2013
Gender	Male	1.07 (0.99 – 1.15)	1.04 (0.95 – 1.14)
	Female	ref	ref
Age	Under 50	0.97 (0.81 – 1.17)	1.04 (0.81 – 1.32)
	50 – 59	0.97 (0.85 – 1.10)	0.96 (0.82 – 1.13)
	60 – 69	ref	ref
	70 – 79	1.04 (0.94 – 1.15)	1.04 (0.92 – 1.19)
	80 and over	0.90 (0.81 – 0.99)	0.95 (0.83 – 1.08)
IMD	1	ref	ref
	2	0.84 (0.70 – 0.99)*	1.02 (0.85 – 1.21)
	3	0.84 (0.70 – 0.99)**	0.84 (0.70 – 1.01)
	4	0.78 (0.65 – 0.94)***	0.81 (0.66 – 0.99)*
	5	0.75 (0.60 – 0.93)***	0.83 (0.65 – 1.06)
Charlson score	0	ref	ref
	1	0.74 (0.66 – 0.83)***	0.74 (0.65 – 0.85)***
	more than 2	0.64 (0.56 – 0.73)***	0.63 (0.53 – 0.73)***
Stage	I	1.44 (1.24 – 1.67)***	1.34 (1.15 – 1.69)***
	II	1.32 (1.16 – 1.51)***	1.28 (1.10 – 1.49)***
	III	1.30 (1.15 – 1.47)***	1.21 (1.04 – 1.41)***
	IV	ref	ref
	Missing	0.98 (0.84 – 1.15)	0.90 (0.74 – 1.09)
Site	Colon	0.69 (0.63 – 0.75)***	0.72 (0.64 – 0.80)***
	Rectum	ref	ref
Year	2000 – 02	0.67 (0.56 – 0.79)***	-
	2003 – 05	0.77 (0.68 – 0.88)***	-
	2006 – 08	ref	ref
	2009 – 11	1.11 (0.97 – 1.27)	1.10 (0.96 – 1.26)
	2012 - 13	1.17 (1.01 – 1.37)	1.16 (0.99 – 1.35)

4.5.3.2 Presenting symptom leading up to referral

The proportion of patients was significantly different between colon and rectal cancer for all symptoms (at $p < 0.05$), except 'other bowel function' (table 29). As expected a significantly higher proportion of rectal cancer patients presented with rectal bleeding than colon cancer patients (41.1% vs 20.4%). The proportion presenting with a change in bowel habit (17% vs 10.2%), diarrhoea (14.2% vs 9.9%) or constipation (8.3% vs 6.8%) was also significantly higher in rectal cancer group. In contrast, colon cancer patients were more likely to present with anaemia (18.6% vs 4.2%), abdominal pain (23.3% vs 8.1%) and weight loss or fatigue (7.6% vs 4%). Overall 4729 (66.3%) patients presented a red-flag

symptom, with the remaining only having a non-specific symptom (i.e. abdominal pain, constipation, other bowel function, weight loss or fatigue). The proportion of patients with a non-specific symptom only was higher in the colon cancer group (39.8% vs 23.0%).

Table 29 - Symptom at first consultation in the year leading up to referral before a colon or rectal cancer diagnosis

	Colon		Rectal		All	
	n	%	n	%	n	%
Rectal bleeding	924 ^a	20.4	1069 ^b	41.1	1993	28.0
Abdominal mass	48 ^a	1.1	14 ^b	0.5	62	0.9
Anaemia	844 ^a	18.6	110 ^b	4.2	954	13.4
Change in bowel habit	461 ^a	10.2	442 ^b	17	903	12.7
Diarrhoea	448 ^a	9.9	369 ^b	14.2	817	11.5
Abdominal pain	1053 ^a	23.3	212 ^b	8.1	1265	17.7
Constipation	307 ^a	6.8	216 ^b	8.3	523	7.3
Other bowel function	96 ^a	2.1	67 ^a	2.6	163	2.3
Weight loss or fatigue	346 ^a	7.6	104 ^b	4	450	6.3
Red-flag	2725 ^a	60.2	2004 ^b	77.0	4729	66.3
Non-specific only	1802 ^a	39.8	599 ^b	23.0	2401	33.7

Superscripts represent differences at $p < 0.05$ between colon and rectal cancer for that particular symptom

4.5.3.3 Delays in referral according to patient demographics

For colon cancer, only patients aged under 50 were significantly more likely than 60 – 69 year olds to experience a delay in referral of between two weeks to three months (OR 1.48 1.02 – 2.14 $p = 0.037$). IMD, gender, year of diagnosis or Charlson score made no significant difference to the odds ratio of a referral between referral interval I and II. However all age groups, except those aged 50 – 59, were significantly more likely than 60 -69 year olds to experience a delay of over three months (Over 80 OR 1.39 95% CI 1.13 – 1.71 $p = 0.002$; 70 – 79 OR 1.23 95% CI 1.01 – 1.51 $p = 0.038$; under 50 OR 1.85 95% CI 1.27 – 2.67 $p = 0.001$). Patients from IMD 5 (OR 1.43 95% CI 1.11 – 1.85 $p = 0.006$), Charlson score one (OR 1.38 95% CI 1.12 – 1.70 $p = 0.003$) and Charlson score two or more (OR 1.38 95% CI 1.18 – 1.70) were also more likely to be referred after three months, relative to the relevant reference categories. Male patients were 29.4% (95% CI 0.61 – 0.82 $p < 0.0005$) less likely to

experience a delay in referral of over three months. Again no difference was seen with the year of diagnosis.

For rectal cancer only patients of IMD 5 were more likely to experience a delay of between two weeks and three months (OR 1.67 95% CI 1.17 – 2.38 $p = 0.004$). Patients aged 80 or over were significantly more likely to be referred after three months (OR 1.49 95% CI 1.18 – 2.35 $p = 0.002$). As with colon cancer, male patients were less likely to experience a delay of over three months (OR 0.71 95% CI 0.59 – 0.87 $p = 0.001$).

Compared with colon cancer patients presenting with rectal bleeding, all other symptoms, except an abdominal mass or change in bowel habit, were associated with delay in referral of between two weeks to three months or over three months. Patients with abdominal mass experienced no significant difference interval to referral. Presenting with a change in bowel habit was likely to lead to a referral after three months than those presenting with rectal bleeding (OR 0.56 95%CI 0.36 – 0.87 $p = 0.01$). Rectal cancer patients presenting with change in bowel habit or an abdominal mass were the equally as likely not be referred late as those with rectal bleeding. All other symptoms were associated with a delay in referral of either two weeks to three months or over three months.

Patients were then restricted according to the presenting symptom, so patients with similar symptoms could be compared. Colon cancer patients with rectal bleeding and a Charlson score of 2 or more were significantly more likely than those with a score of zero to be referred after three months (OR 2.67 95% CI 1.32 – 5.39 $p = 0.006$). Colon cancer patients with a Charlson score of 2 or more presenting with anaemia were significantly more likely to be referred after 3 months (OR 3.26 95% CI 1.98 – 5.35 $p < 0.005$) than patients with a Charlson score of zero with similar symptoms. Female colon cancer patients with this symptom were also more likely to be referred after three months (Male OR 0.65 95% CI 0.46 – 0.92 $p = 0.016$). Colon cancer patients diagnosed in 2000 – 02 or between 2003 – 05 also were more likely to experience a delay in referral of over three months if they presented with anaemia than those diagnosed between 2006 – 08.

Female rectal cancer patients with rectal bleeding were more likely to be referred after three months compared to male patients with the same symptoms (male OR 0.62 95% CI 0.41 – 0.95 $p = 0.027$). No other significant findings were noted for the other variables for any of the presenting symptoms. For the non-specific symptoms this may be due to the relatively low number of patients.

4.5.3.4 Stage at diagnosis and hazard ratio of death according to presenting symptom

Patients presenting with rectal bleeding had the highest proportion of stage I disease for both sites (table 30) colon - 21.3%; rectal - 21.5%). For colon cancer, the non-specific symptoms were associated with a higher stage IV cancer. For example 17.9% of colon cancer patients with abdominal pain were diagnosed with stage IV cancer as were 17.3% of patients with weight loss or fatigue, which were significantly higher than the proportion of patients with stage I or II disease. For rectal cancer, although 'other bowel function' patients demonstrated the highest proportion of stage IV at 14.9%, this did not reach significance. Consequently, red-flag symptoms as a group had a higher proportion of stage I disease for both sites. Overall 18.0% of colon cancer and 24.9% of rectal patients had missing stage data, with abdominal mass the highest for colon (39.6% $p < 0.05$) and constipation for rectal cancer (31.5% $p = NS$).

Table 30 Stage at diagnosis for each presenting symptom, stratified according to site of cancer

Presenting symptom	Colon cancer Stage (% per symptom)					Rectal Cancer Stage (% per symptom)				
	I	II	III	IV	Missing	I	II	III	IV	Missing
Red-flag	14.4 ^a	31.3 ^b	27.0 ^b	9.8 ^c	17.5 ^b	17.3 ^a	21.5 ^a	27.5 ^a	10.5 ^{a,b}	23.3 ^b
Non-specific	6.4 ^a	31.0 ^b	27.2 ^b	16.5 ^c	18.8 ^b	14.9 ^a	18.7 ^a	23.9 ^a	12.4 ^{a,b}	30.2 ^b
Rectal bleeding	21.3 ^a	26.2 ^b	26.9 ^c	8.0 ^d	17.5 ^{b,c}	21.5 ^a	19.5 ^{b,c}	28.5 ^c	8.4 ^d	22.1 ^{b,d}
Anaemia	9.8 ^{a,b}	37.0 ^c	28.4 ^c	9.0 ^a	15.8 ^{a,b}	17.3 ^{a,b,c}	30.9 ^c	17.3 ^b	6.4 ^{a,b}	28.2 ^{a,c}
Change in bowel habit	11.5 ^a	30.2 ^a	27.8 ^a	11.3 ^a	19.3 ^a	15.6 ^{a,b}	23.1 ^b	27.8 ^b	13.3 ^b	20.1 ^a
Diarrhoea	12.5 ^a	32.6	25.7	12.9	16.3	10.3 ^a	19.5 ^b	26.8 ^{b,c}	14.4 ^c	29.0 ^{a,c}
Abdominal mass	6.3 ^{a,b,c}	29.2 ^c	8.3 ^b	16.7 ^{c,d}	39.6 ^d	7.1 ^a	28.6 ^a	35.7 ^a	7.1 ^a	21.4 ^a
Abdominal pain	6.5 ^a	30.7 ^{b,c}	28.8 ^c	17.9 ^d	16.1 ^b	10.8 ^a	21.7 ^{a,b}	28.3 ^b	10.4 ^{a,b}	28.8 ^b
Constipation	6.8 ^a	30.3 ^b	28.3 ^b	12.1 ^{a,b}	22.5 ^b	15.7 ^{a,b}	15.7 ^b	22.7 ^b	14.4 ^a	31.5 ^a
Other bowel function	7.3 ^a	37.5 ^a	24.0 ^a	12.5 ^a	18.8 ^a	11.9 ^a	22.4 ^a	20.9 ^a	14.9 ^a	29.9 ^a
Weight loss or fatigue	5.8 ^a	30.9 ^{b,c}	22.5 ^c	17.3 ^d	23.4 ^{b,d}	23.1 ^a	16.3 ^{a,b}	19.2 ^b	10.6 ^{a,b}	30.8 ^{a,b}
All	11.2	31.2	27.1	12.5	18.0	16.7	20.9	26.7	10.9	24.9

Different superscripts indicate a significant difference in stage for that particular symptom at the $p < 0.05$ level.

All presenting symptoms were associated with a worse prognosis compared with per rectal bleeding, which reflected the lower proportion of stage III or IV in the latter (table 31). For example, the HR for colon cancer patients with abdominal pain was 2.00 (1.71 – 2.34) (stage III - 28.8% and stage IV - 17.9%) and for anaemia was 1.40 (1.18 – 1.67) (stage III 28.4% and stage IV 9.0%) compared with 26.9% stage III and 8.0% Stage IV for rectal bleeding (table 31). A general trend of an association between a higher proportion of stage IV disease and raised HR was noted. The highest significant HR relative to rectal bleeding was abdominal mass for colon cancer (HR 3.99 95% CI 2.75 – 5.78 $p < 0.0005$) and constipation for rectal cancer (HR 1.87 95% CI 1.49 – 2.34 $p < 0.005$). When grouped together, colon or rectal patients with non-specific symptoms had a worse outcome than those with red-flag symptoms (HR 1.49 (95% CI 1.36 – 1.64) and 1.84 (1.59 – 2.14) respectively).

After excluding patients who would have experienced a delay in referral of over two weeks, the HR relative to rectal bleeding increased for all symptoms, except for 'other bowel function', which no longer was significant for both cancers. Of note anaemia and abdominal mass for rectal cancer became significant after the cohort was restricted to those referred within two weeks of the first consultation.

Table 31 Hazard ratio (HR) of death for colon and rectal cancer patients according to presenting symptom compared with either red-flag symptoms as a group or rectal bleeding only

Presenting symptom	Colon cancer		Rectal cancer	
	All patients	Referral interval I only	All patients	Referral interval I only
Red-flag	ref	ref	ref	ref
Non-specific	1.49 (1.36 – 1.64)***	1.84 (1.59 – 2.14)***	1.45 (1.26 – 1.67)***	1.81 (1.42 – 2.31)***
Rectal bleeding	Ref	Ref	ref	ref
Anaemia	1.40 (1.18 – 1.67)***	1.36 (1.08 – 1.70)**	1.30 (0.94 – 1.78)	1.92 (1.28 – 2.87)***
Change in bowel habit	1.33 (1.08 – 1.62)***	1.48 (1.16 – 1.88)**	1.34 (1.11 – 1.61)**	1.53 (1.25 – 1.89)***
Diarrhoea	1.72 (1.41 – 2.08)***	1.95 (1.47 – 2.58)***	1.68 (1.39 – 2.03)***	1.63 (1.23 – 2.15)***
Abdominal mass	3.99 (2.75 – 5.78)***	4.60 (3.05 – 6.95)***	1.92 (0.90 – 4.06)	3.10 (1.24 – 7.30)***
Abdominal pain	2.00 (1.71 – 2.34)***	2.51 (2.03 – 3.11)***	1.53 (1.21 – 1.93)***	1.83 (1.26 – 2.66)***
Constipation	1.70 (1.36 – 2.11)***	1.90 (1.28 – 2.83)***	1.87 (1.49 – 2.34)**	2.61 (1.73 – 3.94)***
Other bowel function	1.45 (1.01 – 2.09)*	1.67 (0.68 – 4.08)	1.61 (1.10 – 2.38)*	1.74 (0.72 – 4.22)
Weight loss or fatigue	2.38 (1.95 – 2.90)***	2.76 (2.05 – 3.72)***	1.69 (1.24 – 2.29)**	3.55 (2.07 – 6.09)***

HR has been adjusted for gender, age, Charlson score and IMD level. * 0.005 - < 0.05 ** 0.0005 - < 0.005 *** < 0.0005

4.5.3.5 Intervals along the patient pathway

Figures 25 – 27

For all colorectal cancer patients, the median primary care interval was 9 days (IQR 0 – 77), the hospital interval was 42 days (IQR 20 – 83) and the diagnostic interval was 78 days (IQR 34 – 181). For both colon and rectal cancer patients presenting with rectal bleeding, change in bowel habit or those noted to have an abdominal mass on examination were referred within a median of one day. All other symptoms were associated with a significantly longer primary care interval compared with rectal bleeding, in particular constipation for colon cancer (median 75 days IQR 16 - 209) and weight loss/fatigue for rectal cancer (median 100 days IQR 21.25 – 213.5).

For colon cancer, only a change in bowel habit (median 41 day IQR 22 - 75) or an abdominal mass (median 30.5 days IQR 15 – 53.5) at presentation significantly decreased the hospital interval compared with rectal bleeding (median 50 days IQR 23 – 91). For rectal cancer, no symptom was associated with a decrease in the hospital interval compared with rectal bleeding (median 29 days IQR 14 – 62). Colon cancer patients with anaemia (median 55 days IQR 28 - 99) or rectal cancer patients with abdominal pain (median 42.5 days IQR 17 - 88) or weight loss/fatigue (median 38 days IQR 19.5 – 83.75) demonstrated a significantly increased hospital interval compared with rectal bleeding.

Colon cancer patients with non-specific symptoms as a group had a significantly longer referral (median 42 days IQR 8 – 160.25 vs 4 days IQR 0 - 40) and diagnostic interval (median 123 days IQR 59 – 259 vs 73 IQR 35 – 155) compared with those with red-flag symptoms, although no difference was seen in the hospital interval. For rectal cancer, patients with non-specific symptoms experienced a significantly longer referral, hospital and diagnostic interval.

Figure 25 - Primary care interval (top left panel), hospital interval (top right panel) and diagnostic interval (bottom panel) in days for colon patients according to presenting symptom

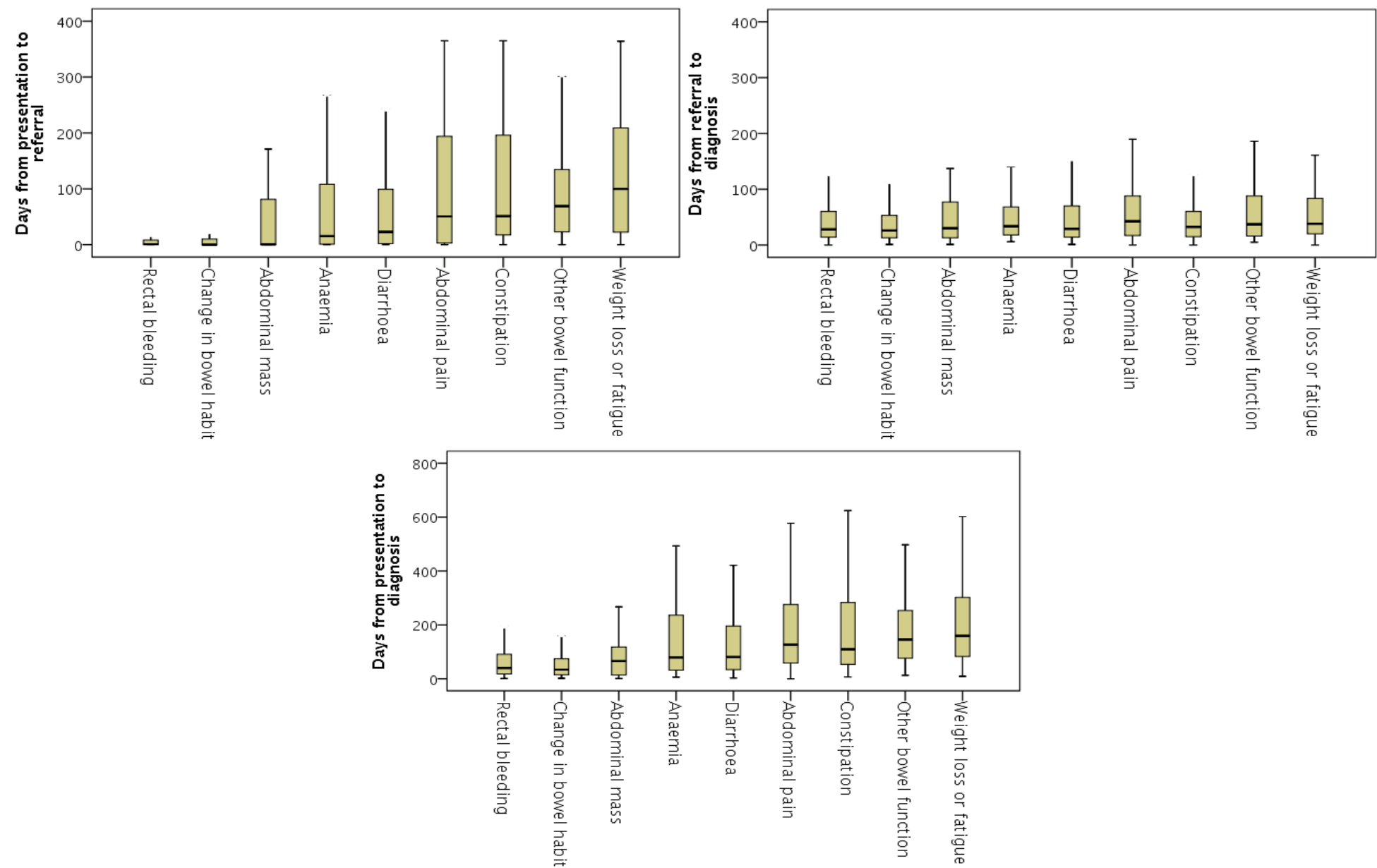


Figure 26 Primary care interval (top left panel), hospital interval (top right panel) and diagnostic interval (bottom panel) in days for rectal patients according to presenting symptom

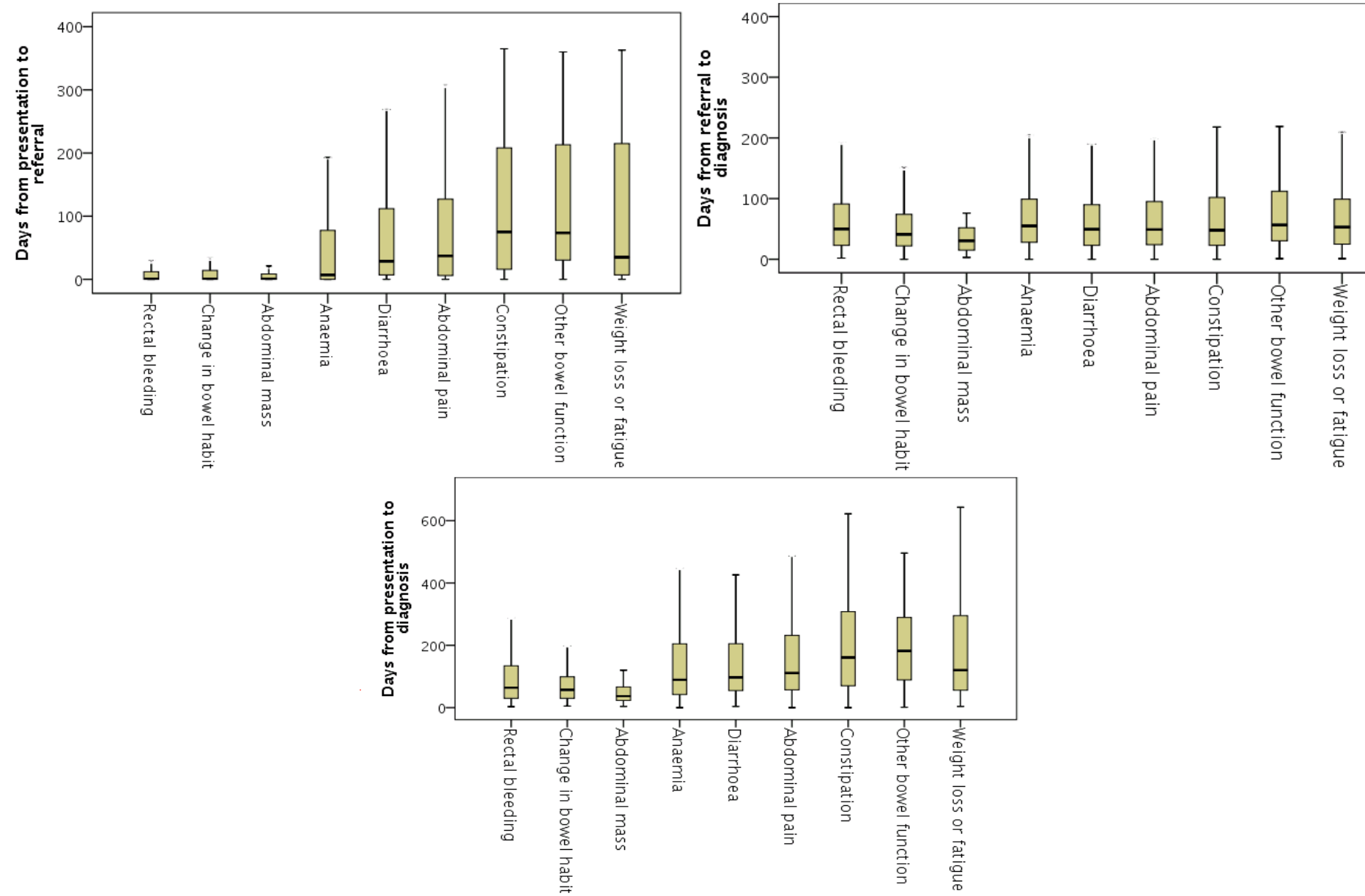
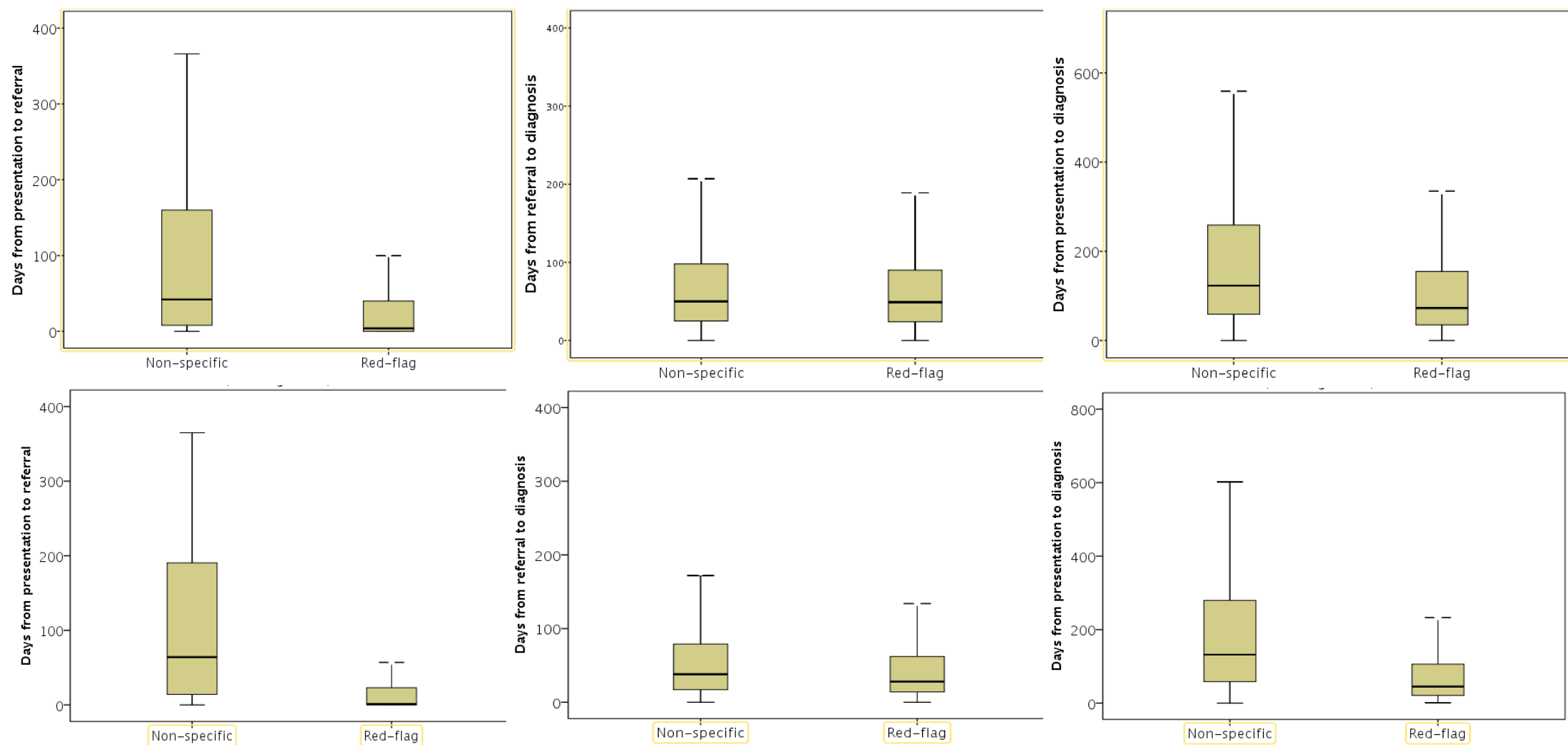


Figure 27 - Primary care interval (left), hospital interval (middle) and diagnostic interval (right) for colon (upper row) and rectal (lower row) cancer patients according to the presence of a red-flag symptom at presentation



4.5.3.6 Change in hospital interval following a delay in referral

For colon cancer, a referral delay of between two week and three months did not significantly delay the hospital interval compared with referrals within two weeks (table 32). However patients who were referred after three months demonstrated a significantly longer hospital interval compared with those in Group I if rectal bleeding, anaemia, abdominal pain or constipation were the presenting symptoms (median increase of 19.5 days, 11.5 days, 14 days and 20 days respectively). For rectal cancer only delays of over three months for a change in bowel habit or diarrhoea significantly increased the hospital interval by a median of 22 and 15 days respectively (table 33).

4.5.3.7 Delays in referral and prognosis

Overall a delay in referral was associated with a worse prognosis for interval groups II (HR 1.28 (1.14 – 1.44)) and III (HR 1.31 (1.16 – 1.47)) for colon cancer (table 32) and only interval II (HR 1.33 (1.13 – 1.55)) for rectal cancer patients (table 33).

Table 32 - adjusted hazard ratio of death for colon cancer patients according to interval to referral

Symptom	Interval	Number (%)	Median hospital interval (IQR)	HR
Red-flag	I	1739 (63.8)	45 (23 – 85)	ref
	II	526 (19.3)	49.5 (25 – 85)	1.28 (1.09 – 1.51)**
	III	460 (16.9)	60.5 (28 – 118.5)***	1.53(1.29 – 1.81)***
Non-specific	I	582 (32.3)	44 (22 – 85.25)	ref
	II	576 (32.0)	49 (25 – 89)	0.94 (0.79 – 1.12)
	III	644 (35.7)	57 (28 – 113)***	0.85 (0.72 – 1.01)
Rectal bleeding	I	704 (76.2)	45 (22 – 90.75)	ref
	II	116 (12.6)	51.5 (24.5 – 91.5)	1.34 (0.93 – 1.93)
	III	104 (11.3)	64.5 (37 – 104.25)**	1.84 (1.29 – 2.63)**
Anaemia	I	488 (57.8)	49.5 (26.25 – 90)	ref
	II	259 (18.8)	61 (33 – 98)	1.27 (1.05 – 1.72)*
	III	197 (23.3)	62 (30.5 – 128)*	1.70 (1.30 – 2.23)***
Change in bowel habit	I	347 (75.3)	42 (22 – 77)	ref
	II	87 (18.9)	38 (24 – 71)	1.14 (0.77 – 1.70)
	III	27 (5.9)	38 (23 – 83)	0.99 (0.48 – 2.06)
Diarrhoea	I	161 (35.9)	43 (22 – 83.5)	ref
	II	159 (35.5)	48 (23 – 82)	1.01 (0.71 – 1.42)
	III	128 (28.6)	57 (24.25 – 134.75)	1.06 (0.73 – 1.53)
Abdominal pain	I	369 (35.0)	43 (21.5 – 83)	ref
	II	358 (34.0)	49 (25 – 87.5)	0.92 (0.74 – 1.15)
	III	326 (31.0)	57 (27.75 – 115.5)**	0.76 (0.59 – 0.95)*
Constipation	I	74 (24.1)	35 (15 – 87.5)	ref
	II	90 (29.3)	46.5 (22 – 89)	1.01 (0.61 – 1.65)
	III	143 (46.6)	55 (28 – 129)*	0.97 (0.62 – 1.53)
Weight loss or fatigue	I	125 (36.1)	49 (22.5 – 91.5)	ref
	II	92 (26.6)	56 (27 – 100.75)	0.98 (0.67 – 1.45)
	III	129 (37.3)	61 (25.5 – 104.5)	0.97 (0.67 – 1.41)
All patients	I	2321 (51.3)	45 (23 – 85)	ref
	II	1102 (24.3)	49 (25 – 87)	1.28 (1.14 – 1.44)***
	III	1104 (24.4)	58.5 (28 – 114.75)***	1.31 (1.16 – 1.47)***

Interval i – less than 2 weeks, interval ii – 2 weeks to 3 months, interval iii – over 3 months. Hazard ratio (HR) adjusted for age, gender, Charlson score, IMD, and interval from referral to diagnosis * 0.005 - < 0.05 ** 0.0005 - < 0.005 *** < 0.0005 data for 'other bowel function' and 'abdominal mass' not shown as too few cases of each to carry out analysis

Table 33 - Adjusted hazard ratio of death for rectal cancer patients according to interval to referral

Symptom	Interval	Number (%)	Median hospital interval (IQR)	HR
Red-flag	I	1414 (70.6)	27 (14 – 58,25)	ref
	II	319 (15.9)	27 (14 – 59)	1.11 (0.90 – 1.36)
	III	271 (13.5)	40 (20 – 78)***	1.30 (1.06 – 1.60)*
Non-specific	I	153 (25.5)	29 (14 – 77.5)	ref
	II	189 (31.6)	37 (18.5 – 61)	1.14 (0.85 – 1.54)
	III	257 (42.9)	43 (17.5 – 87)***	0.57 (0.42 – 0.78)***
Rectal bleeding	I	848 (79.3)	28 (14 – 61)	ref
	II	121 (11.3)	29 (15.5 – 57.5)	1.01 (0.71 – 1.45)
	III	100 (9.4)	36.5 (19.25 – 69.75)	1.65 (1.18 – 2.31)**
Anaemia	I	54 (49.1)	32.5 (18 – 61)	ref
	II	26 (23.6)	27.5 (14 – 60)	0.73 (0.33 – 1.63)
	III	30 (27.3)	44.5 (15.25 – 91.5)	0.77 (0.34 – 1.76)
Change in bowel habit	I	348 (78.7)	24 (13 – 49)	ref
	II	57 (12.9)	26 (13.5 – 58)	0.79 (0.49 – 1.29)
	III	37 (8.4)	46 (26 – 106)**	0.78 (0.41 – 1.72)
Diarrhoea	I	155 (42.0)	27 (14 – 63)	ref
	II	113 (30.6)	24 (13 – 65)	1.30 (0.90 – 1.89)
	III	101 (27.4)	42 (21 – 80)*	1.13 (0.77 – 1.67)
Abdominal pain	I	68 (32.1)	42.5 (16.25 – 84)	ref
	II	61 (28.8)	41 (18 – 83.5)	1.19 (0.72 – 1.98)
	III	83 (39.2)	49 (17 – 107)	0.57 (0.33 – 1.00)*
Constipation	I	51 (23.6)	42.5 (16.25 – 84)	ref
	II	75 (34.7)	41 (18 – 83.5)	1.29 (0.77 – 2.14)
	III	90 (41.7)	49 (17 – 107)	0.68 (0.40 – 1.17)
Weight loss or fatigue	I	22 (21.2)	32 (16.25 – 85)	ref
	II	26 (25.0)	37 (16.5 – 73.5)	1.05 (0.45 – 2.41)
	III	56 (53.8)	43.5 (21.25 – 93.25)	0.37 (0.18 – 0.78)**
All patients	I	1567 (60.2)	27 (14 – 60)	ref
	II	508 (19.5)	31 (14 – 59.75)	1.33 (1.13 – 1.55)***
	III	528 (20.3)	42 (19 – 84)***	1.08 (0.92 – 1.27)

Interval i – less than 2 weeks, interval ii – 2 weeks to 3 months, interval iii – over 3 months. Hazard ratio (HR) adjusted for age, gender, Charlson score, IMD, and interval from referral to diagnosis * 0.005 - < 0.05 ** 0.0005 - < 0.005 *** < 0.0005 data for 'other bowel function' and 'abdominal mass' not shown as too few cases of each to carry out analysis

Colon cancer patients with rectal bleeding who experienced a delay in primary care of over three months (group III - HR 1.84 (1.29 – 2.63)) demonstrated a significantly worse prognosis compared with patients with rectal bleeding referred within two weeks (table 32). This corresponds to an increase in stage III and IV cancer in interval group III (table 34). Although a similar trend was seen for patients in interval group II, this did not reach significance. Colon cancer patients with red-flag symptoms demonstrated an increase in HR for both interval group II (HR 1.28 (1.09 – 1.51)) and III (HR 1.53 (1.29 – 1.81)), with a concomitant rise in late stage at diagnosis (from 8.2% stage IV in interval group I, 11.4% in group II to 14.1% in group III). A similar finding was noted in the sub-group analysis of colon cancer patients with anaemia.

Rectal cancer patients with rectal bleeding also had a worse prognosis (HR 1.65 (1.18 – 2.31)) if referral was delayed to over three months (table 33), which corresponded to a decrease in stage I from 22.1% to 12.0% and an increase in missing stage from 21.8% to 26.0%, even though the overall difference in stage did not reach significance (table 35). Again patients with red-flag symptoms and a delay of over three months had a worse prognosis (HR 1.30 (1.06 – 1.60)) compared with those referred within two weeks.

Both colon and rectal cancer patients with abdominal pain (HR 0.76 (95% CI 0.59 – 0.95) and HR 0.57 (0.33 – 1.00) respectively) and rectal cancer patients with weight loss/fatigue (HR 0.37 (95% CI 0.18 – 0.78)) or non-specific symptoms (HR 0.57 (95% CI 0.42 – 0.78)) as a group demonstrated an improved prognosis if referral was delayed to beyond three months, which is reflected in the improved stage at diagnosis. There was no difference in outcome if either colon or rectal cancer patients had experienced a delay with the other symptoms.

Table 34 - Stage at diagnosis for colon cancer stratified according to symptom at presentation

Symptom	Interval	Number (%)	Stage					P value
			I	II	III	IV	Missing	
Red-flag	I	1739 (63.8)	15.0	31.4	28.2	8.2	17.2	0.014
	II	526 (19.3)	13.3	31.9	25.9	11.4	17.5	
	III	460 (16.9)	13.3	30.2	23.9	14.1	18.5	
Non-specific	I	582 (32.3)	7.0	28.5	26.8	16.4	20.3	0.592
	II	576 (32.0)	5.7	31.6	28.3	17.5	16.8	
	III	644 (35.7)	6.5	32.8	26.7	14.9	19.1	
Rectal bleeding	I	704 (76.2)	21.9	26.4	27.0	6.7	18.0	0.017
	II	116 (12.6)	22.4	27.6	25.0	6.9	18.1	
	III	104 (11.3)	16.3	23.1	28.8	18.3	13.5	
Anaemia	I	488 (57.8)	10.0	39.3	31.4	6.4	12.9	0.004
	II	259 (18.8)	8.8	37.7	22.6	12.6	18.2	
	III	197 (23.3)	10.2	30.5	25.9	12.7	20.8	
Change in bowel habit	I	347 (75.3)	11.2	30.5	28.8	10.4	19.0	0.695
	II	87 (18.9)	13.8	28.7	24.1	16.1	17.2	
	III	27 (5.9)	7.4	29.6	25.9	7.4	29.6	
Diarrhoea	I	161 (35.9)	9.9	30.4	28.0	13.7	18.0	0.119
	II	159 (35.5)	11.3	31.4	30.4	11.3	14.5	
	III	128 (28.6)	17.2	36.7	15.6	14.1	16.4	
Abdominal pain	I	369 (35.0)	8.4	27.9	27.4	18.2	18.2	0.146
	II	358 (34.0)	4.5	32.7	39.9	20.1	12.8	
	III	326 (31.0)	6.4	31.6	29.1	15.3	17.5	
Constipation	I	74 (24.1)	6.8	23.0	32.4	16.2	21.6	0.836
	II	90 (29.3)	6.7	33.3	27.8	8.9	23.3	
	III	143 (46.6)	7.0	32.2	26.6	11.9	22.4	
Weight loss or fatigue	I	125 (36.1)	4.0	33.6	20.0	16.8	25.6	0.597
	II	92 (26.6)	8.7	25.0	22.8	16.3	27.2	
	III	129 (37.3)	5.4	32.6	24.8	18.6	18.6	
All patients	I	2321 (51.3)	13.0	30.7	27.8	10.5	18.0	< 0.0005
	II	1102 (24.3)	9.3	31.8	27.1	14.6	17.2	
	III	1104 (24.4)	9.3	31.7	25.5	14.6	18.8	

Interval i – less than 2 weeks, interval ii – 2 weeks to 3 months, interval iii – over 3 months. Data for ‘other bowel function’ and ‘abdominal mass’ not shown as too few cases of each to carry out analysis

Table 35 - Stage at diagnosis for rectal cancer stratified according to presenting symptom

Symptom	Interval	Number (%)	Stage					P value
			I	II	III	IV	Missing	
Red-flag	I	1414 (70.6)	19.0	21.4	27.1	10.5	21.9	0.05
	II	319 (15.9)	13.8	22.9	28.2	11.0	24.1	
	III	271 (13.5)	12.2	20.3	28.8	9.6	29.2	
Non-specific	I	153 (25.5)	11.8	22.2	20.9	13.7	31.4	0.330
	II	189 (31.6)	12.7	18.0	26.5	14.8	28.0	
	III	257 (42.9)	18.3	17.1	23.7	9.7	31.1	
Rectal bleeding	I	848 (79.3)	22.1	19.6	28.5	8.0	21.8	0.269
	II	121 (11.3)	16.5	22.3	28.9	11.6	20.7	
	III	100 (9.4)	12.0	26.0	28.0	8.0	26.0	
Anaemia	I	54 (49.1)	22.2	29.6	16.7	3.7	27.8	0.170
	II	26 (23.6)	11.5	46.2	3.8	7.7	30.8	
	III	30 (27.3)	13.3	20.0	30.0	10.0	26.7	
Change in bowel habit	I	348 (78.7)	15.8	23.3	26.4	14.1	20.4	0.965
	II	57 (12.9)	14.0	22.8	35.1	10.5	17.5	
	III	37 (8.4)	16.2	21.6	29.7	10.8	21.6	
Diarrhoea	I	155 (42.0)	9.0	24.5	23.9	18.7	23.9	0.167
	II	113 (30.6)	11.5	17.7	29.2	11.5	30.1	
	III	101 (27.4)	10.9	13.9	28.7	10.9	35.6	
Abdominal pain	I	68 (32.1)	10.3	26.5	22.1	8.8	32.4	0.782
	II	61 (28.8)	9.8	18.0	31.1	14.8	26.2	
	III	83 (39.2)	12.0	20.5	31.3	8.4	27.7	
Constipation	I	51 (23.6)	15.7	17.6	23.5	13.7	29.4	0.421
	II	75 (34.7)	13.3	22.7	20.0	17.3	26.7	
	III	90 (41.7)	17.8	8.9	24.4	12.2	36.7	
Weight loss or fatigue	I	22 (21.2)	13.6	13.6	22.7	22.7	27.3	0.220
	II	26 (25.0)	26.9	3.8	23.1	11.5	34.6	
	III	56 (53.8)	15.2	18.8	26.3	9.7	20.1	
All patients	I	1567 (60.2)	18.3	21.5	26.5	10.8	22.8	0.018
	II	508 (19.5)	13.4	21.1	27.6	12.4	25.6	
	III	528 (20.3)	15.2	18.8	26.3	9.7	30.1	

*interval i – less than 2 weeks, interval ii – 2 weeks to 3 months, interval iii – over 3 months. Hazard ratio (hr) adjusted for age, gender, charlson score, imd, and interval from referral to diagnosis * 0.005 - < 0.05 ** 0.0005 - < 0.005 *** < 0.0005 data for 'other bowel function' and 'abdominal mass' not shown as too few cases of each to carry out analysis*

4.5.3.8 Number of consultations before referral

In total 1709 (37.8%) of colon cancer and 1266 (48.6%) of rectal cancer patients were referred after the first consultation, with 1811 (40.0%) of colon and 833 (32.0%) of rectal cancer patients referred after three or more consultations (table 36 and 37).

The red-flag symptoms analysed individually demonstrated the highest percentage of patients referred after just one consultation, and subsequently the lowest percentage referred after three or more consultations. The percentage of patients referred after one consultation ranged from 66.7% for colon cancer patients with an abdominal mass to 7.3% of patients with 'Other bowel function'. For rectal cancer the highest percentage of patients referred after one consultation was for rectal bleeding (65.4%), down to 11.9% for 'other bowel function'. The symptoms with the highest proportion of patients referred after three or more consultations corresponded to the symptoms with the longest median interval to referral described in the previous study.

Table 36 - Number of consultations before referral for colon cancer patients

	One		Two		Three or more		p
	n	%	n	%	n	%	
All patients	1709	37.8	1007	22.2	1811	40.0	< 0.0005
Red-flag	1302 ^a	47.8	575 ^b	21.1	848 ^c	31.1	
Non-specific only	407 ^a	22.6	432 ^b	24.0	963 ^c	53.4	
Rectal bleeding	565 ^a	61.1	185 ^b	20.0	174 ^c	18.8	< 0.0005
Change in bowel habit	258 ^a	56.0	108 ^b	23.4	95 ^c	20.6	
Abdominal mass	32 ^a	66.7	12 ^a	25.0	4 ^b	8.3	
Anaemia	349 ^a	41.4	183 ^{a,b}	21.7	312 ^b	37.0	
Diarrhoea	98 ^a	21.9	87 ^b	19.4	263 ^b	58.7	
Abdominal pain	260 ^a	24.7	252 ^b	23.9	541 ^b	51.4	
Constipation	63 ^a	20.5	65 ^b	21.2	179 ^c	58.3	
Other bowel function	7 ^a	7.3	18 ^b	18.8	71 ^b	74.0	
Weight loss or fatigue	77 ^a	22.3	97 ^b	28.0	172 ^b	49.7	

Table 37 - Number of consultations before referral for rectal cancer patients

	One		Two		Three or more		p
	n	%	n	%	n	%	
All patients	1266	48.6	504	19.4	833	32.0	< 0.0005
Red-flag	1145 ^a	57.1	362 ^b	18.1	497 ^c	24.8	
Non-specific only	121 ^a	20.2	142 ^b	23.7	336 ^c	56.1	
Rectal bleeding	699 ^a	65.4	184 ^b	17.2	186 ^c	17.4	< 0.0005
Change in bowel habit	293 ^a	66.3	72 ^b	16.3	77 ^c	17.4	
Abdominal mass	9 ^a	64.3	1 ^a	7.1	4 ^a	28.6	
Anaemia	38 ^a	34.5	22 ^{a,b}	20.0	50 ^b	45.5	
Diarrhoea	106 ^a	28.7	83 ^b	22.5	180 ^b	48.8	
Abdominal pain	62 ^a	29.2	47 ^b	22.2	103 ^b	48.6	
Constipation	32 ^a	14.8	50 ^b	23.1	134 ^c	62.0	
Other bowel function	8 ^a	11.9	19 ^b	28.4	40 ^b	59.7	
Weight loss or fatigue	19 ^a	18.3	26 ^b	25.0	59 ^b	56.7	

4.5.3.9 Relationship between number of consultations and time to referral

The median time to referral increased with increasing consultations before referral, from 0 (IQR 0 – 1) day for colon cancer patients referred after the first consultation to 68 (IQR 23 – 186) days for patients referred after three or more consultations (table 38). For rectal cancer the median days increased from zero (IQR 0 – 1) to 84 (IQR 28 – 199) days respectively.

Table 38 - Median days, and interquartile range, for colon or rectal cancer patients referred after one, two or three or more consultations with relevant symptoms

	One	Two	Three or more
Colon	0 (0 – 1)	17 (6 – 51)	68 (23 – 186)
Rectal	0 (0 – 1)	15 (5 – 45.75)	84 (28 – 199)

For colon cancer cases, the Spearman's Coefficient was calculated to be 0.89 $p < 0.005$ and for rectal cancer it was 0.90 $p < 0.0005$, indicating a strong correlation between the number of consultations and referral interval for both sites.

The Area Under the Curve (AUC) for one or two consultations versus three or more was 0.89 (95% CI 0.88 – 0.90 $p < 0.0005$) for colon cancer patients and 0.94 (95% CI .93 - .95 $p < 0.0005$) for rectal cancer. The AUC for one, two, or three consultations versus four or more was 0.87 (95% CI 0.86 – 0.88 $p < 0.0005$) for the colon cancer cases and 0.90 (95% CI 0.89 – 0.92 $p < 0.0005$) for rectal cancer cases. This suggests the binary category of one or two versus three or more is a better predictor of the number of days to referral.

4.5.3.10 Factors associated with a referral after three or more consultations

For colon cancer, a Charlson score or more than one (OR 1.26 95% CI 1.03 – 1.55) was associated with an increased association with a referral after three or more consultations. The presence of a red-flag symptom at first consultation (OR 0.39 95% CI 0.34 – 0.44 $p < 0.0005$), or male patients (OR 0.88 95% CI 0.78 – 0.99 $p = 0.045$) were associated with a reduced odds of a delayed referral. For rectal cancer, only the presence of a red-flag symptom reduced the odds of a late referral (OR 0.26 95% CI 0.21 – 0.32 $p < 0.0005$)

4.5.3.11 Association between delays in referral and survival

Colon cancer patients aged 80 and over demonstrated an increased hazards ratio of death compared with those aged 60 – 69 (HR 1.62 95% CI 1.41 - .85 $p < 0.0005$), as did patients from IMD 4 (HR 1.21 95% CI 1.04 – 1.41 $p = 0.012$) and IMD 5 (HR 1.32 95% CI 1.11 – 1.56 $p = 0.001$). In the model without the presenting symptom defined, patients referred after three or more consultations demonstrated a worse prognosis (HR 1.24 95% CI 1.13 - .37 $p < 0.0005$) (see Figure 28 for Kaplan Meier graph).

Rectal cancer patients aged 80 and over (HR 2.32 95% CI 1.92 – 2.81 $p < 0.0005$) or 70 – 79 year olds (HR 1.56 95% CI 1.30 – 1.87 $p < 0.0005$) demonstrated a worse prognosis than those aged 60 – 69. A Charlson score of more than one also worsened prognosis (HR 1.69 95% CI 1.34 – 2.15 $p < 0.0005$). In the model without the presenting symptom defined, patients referred after three or more consultation demonstrated a worse prognosis (HR 1.17 95% CI 1.03 – 1.34 $p = 0.019$) (see Figure 29 for Kaplan Meier graph).

For both colon and rectal cancer patients presenting with rectal bleeding, a delay of three or more consultations was associated with a worse prognosis (HR 1.38 95% CI 1.02 – 1.86 $p = 0.037$ and HR 1.35 95% CI 1.03 – 1.78 $p = 0.029$ respectively) (table 39). In addition, for colon cancer, delays in referral for patients presenting with anaemia (HR 1.33 95% CI 1.05 – 1.68 $p = 0.017$) or diarrhoea (HR 1.39 95% CI 1.03 – 1.89 $p = 0.034$) were also associated with a worse prognosis than patients referred after a maximum of two consultations. There were no other associations that reached the significance level of $p < 0.05$ for either colon or rectal cancer.

Figure 28 - Kaplan Meier demonstrating the difference in survival for colon cancer patients referred after one or two and three or more consultations

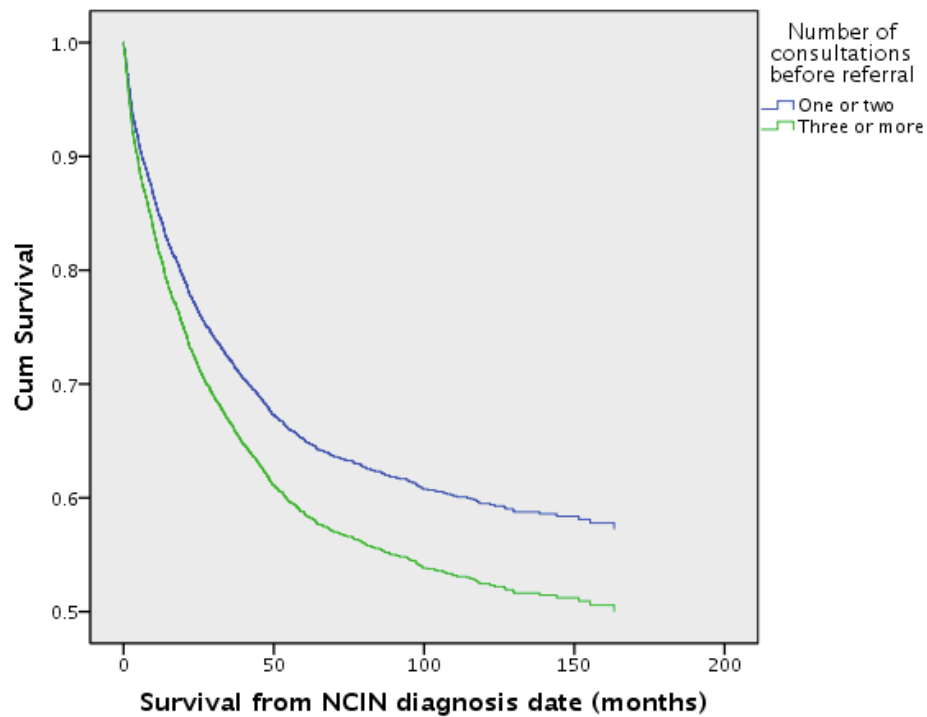


Figure 29 - Kaplan Meier demonstrating the difference in survival for rectal cancer patients referred after one or two and three or more consultations

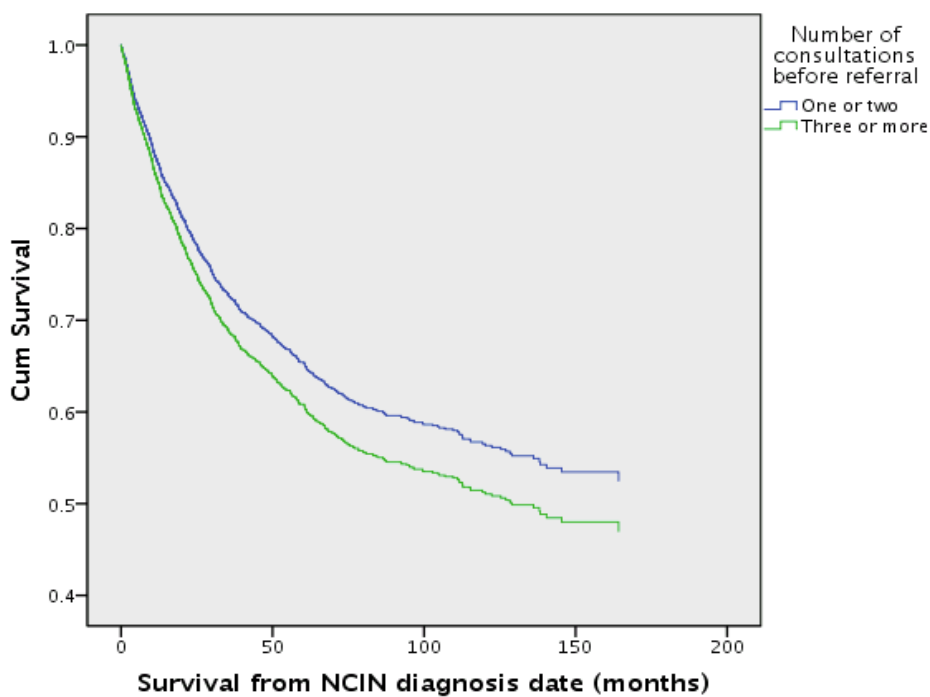


Table 39 - Hazards ratio of death(HR) for a delay in referral (three or more consultations before referral).

	Colon		Rectal	
	HR	p	HR	p
Not defined	1.24 (1.13 – 1.37)	< 0.0005	1.17 (1.03 – 1.34)	0.019
Red-flag	1.39 (1.21 – 1.57)	< 0.0005	1.13 (0.95 – 1.33)	0.166
Non-specific only	0.90 (0.79 – 1.04)	0.157	1.02 (0.80 – 1.31)	0.879
Rectal bleeding	1.38 (1.02 – 1.86)	0.037	1.35 (1.03 – 1.78)	0.029
Change in bowel habit	1.25 (0.86 – 1.82)	0.250	0.84 (0.55 – 1.28)	0.422
Abdominal mass	3.58 (0.73 – 17.5)	0.115	-	
Anaemia	1.33 (1.05 – 1.68)	0.017	0.73 (0.38 – 1.42)	0.355
Diarrhoea	1.39 (1.03 – 1.89)	0.034	0.88 (0.65 – 1.20)	0.413
Abdominal pain	0.92 (0.76 – 1.10)	0.357	1.02 (0.67 – 1.56)	0.93
Constipation	0.91 (0.63 – 1.30)	0.595	1.33 (0.87 – 2.03)	0.194
Other bowel function	0.79 (0.35 – 1.78)	0.571	0.86 (0.38 – 1.94)	0.717
Weight loss or fatigue	0.96 (0.70 – 1.30)	0.769	0.80 (0.43 – 1.50)	0.487

Each result represents a separate model, where only patients with the same presenting symptom were included. All models adjusted for age, Charlson score, IMD and gender. There were insufficient number of rectal cancer patients with an abdominal mass to provide a HR

4.5.4 Discussion

4.5.4.1 Summary

This study demonstrates that a delay in referral, either measured in days or number of consultations, is associated with a worse outcome for patients, in particular for patients with red-flag symptoms at first consultation. In addition, compared with patients who present with rectal bleeding, this study has shown survival is worse for patients who present with other cancer associated symptoms, even after exclusion of patients who experienced a delay in referral of over two weeks. Although a patient may have experienced a referral delay of over three months, the hospital interval was lengthened by a median of only 13 days, indicating the primary care interval has a greater influence on diagnostic delays than the latter. 40% of patients present with non-specific symptoms that fall outside the NICE referral guidance, of which 53.4% are seen three or more times before referral. For patients presenting with red-flag symptoms, this drops to 31.1%.

4.5.4.2 Comparison between delays as time or delays as consultations

The number of consultations before referral were included as an outcome variable as this can be a tangible target for health policy. However, not every patient who has been seen three times before referral has waited the same amount of time before referral. The findings of the ROC curve indicate there is a close relationship but there will inevitably be

variation as expressed by the interquartile range. For this reason the time expired before referral was the main focus, as tumours progress over time rather than consultations.

Both studies demonstrated a reduced delay in referral for primary consultations for male patients for colon and rectal cancers. The first study demonstrated a delay with the extremes end of age, which was not noted in the study with consultations. Both studies demonstrated a delay (over three months in the first study) for patients presenting with red-flag symptoms was associated with a worse survival. If all patients were considered together, a delay of between 2 weeks to three months was associated with a worse survival, for both cancer, but only over three months for colon cancer. In the latter study, for both colon and rectal cancer, a delay of three or more consultations was associated with a worse outcome. The findings that certain symptoms were associated with a reduced hazard of death in the first study was not replicated in the next study. These differences between the studies may be explained by the disparity in the interval to referral between the categories. The median time from presentation to referral for colon cancer patients who experienced three or more consultations was less than the threshold of three months used in the first study. However for rectal cancer the median time was closer at 84 days.

4.5.4.3 Comparison with previous studies

A significant difference in stage or survival between urgent or routine pathway referrals has not been conclusively found^{68,72,162,163}. Many of these studies have insufficient patient numbers and have not investigated whether the absolute time taken to refer impacts prognosis. Our study is the first to carry out such analysis with over 7000 patients. Two national audits carried out since 2009 allow a comparison with the data and methodology utilised here. The National Audit of Cancer diagnosis in primary care was completed through questionnaires sent to the GP, who were advised to use their medical records¹⁶⁴. 21.3% of colorectal cancer patients experienced 3 or more consultations before referral. Referral after one, two or three consultations was equivalent to a median referral interval of 0 (0 – 3.5), 18 (7 – 40) and 39 (15 – 75.5) days. The interval for one and two consultations corresponds well to our findings. The median time to referral for three or more consultations was not available from the audit. From this same audit the primary care interval was shown to be 6 (0 – 29) days for all colorectal cancer patients, compared with 9

(0 – 77) days in this study. From the National Cancer Patient Experience Survey¹⁶⁵, which asked patients, rather than GPs, to complete questionnaires, 32.8% of colon and 22.5% of rectal patients experienced 3 or more consultations. This compares with 40% of colon and 32% of rectal cancer patients in this study. Although the interval for referral for those who experienced less than three consultations in this study is similar to the GP audit, there is a larger discrepancy in the percentage of patients who experienced three or more consultations before referral. However the percentage of patients with three or more consultations was lowest for red-flag symptoms, and higher for non-specific symptoms, corresponding to the differences in the referral interval according to the type of symptom from the first study.

My findings are similar to findings from an observational audit by Barrett et al.⁴⁸ In their study the median interval in days from first presentation to referral was 0 (0 – 1) for abdominal mass, 2 (0 – 7) for rectal bleeding, 33 (0 – 44) for diarrhoea, 45 (0 – 5) for abdominal pain, 7 (1 – 182) for constipation, 11 (0 – 54) for anaemia for colorectal cancer patients. This suggests my definitions of critical time points in the patient pathway correspond to data collected from medical notes and GP questionnaires.

The majority of studies have concentrated on the interval from onset of symptoms to diagnosis or treatment. In a recent meta-analysis, Neal et al.¹⁶⁶ stated it was difficult to make conclusions about the impact of delays on survival for colorectal cancer patients due to differences in study design. Delays leading to worse outcome have been described^{79,167}, although most studies have found no association between delays and outcome^{62,76,81,168,169} or stage^{84,170}. More recent papers have shown short intervals were associated with a worse outcome, presumably due to more aggressive cancers^{75,79,171}. In this study, delays in referral with certain symptoms were associated with a better stage and prognosis, again perhaps reflecting tumour biology.

Patients with red-flag symptoms experienced the shortest interval and least consultations before referral, reflecting the national referral guidelines from NICE^{52,172} and previous studies^{32,36}. My study confirms patients with per rectal bleeding are diagnosed at an earlier stage^{62,173,174} and so demonstrate a better prognosis^{168,174–176}, while abdominal pain¹⁷¹ and

change in bowel habit are related to a later stage^{62,173}. A similar finding was noted after excluding patients who were referred after two weeks, indicating it is the symptom itself, rather than delays in referral, that accounts for the difference in stage and survival.

The hospital interval was less varied between red-flag and non-specific symptoms, as patients referred with colorectal associated symptoms would be expected to be seen in outpatients, undergo either an endoscopic examination and/or radiological investigation, and then await histological confirmation. The median hospital interval was slightly longer for patients who were referred after three months. In such cases the patients were perhaps referred through a non-urgent pathway following a period of 'watch and wait' as the GP did not have a high enough suspicion of cancer. As the primary care interval of over three months was associated with only a modest increase in the hospital interval, the delay in referral is more likely to account for the differences in stage and survival seen in this study.

4.5.4.4 Strengths and limitations

I have been able to include a large number of patients, with a varied set of symptoms, from a data source that has been shown to be representative of the national population¹²⁵. The data is not liable to recall bias, as the clinical information was recorded at the time of consultation. The finding that delays in primary care are associated with a worse prognosis is reflected by a later tumour stage. This supports the theory that delays lead to a progression of disease. This is also the first study to describe the hospital interval for individual colorectal cancer associated symptoms.

With CPRD it is not possible to determine the duration of symptoms before the patient presents to the GP and it is known that the patient interval varies according to symptom³⁶. However I compared the effect of delays for patients with the same presenting symptom to minimize the impact of this missing data. Symptom combinations may have different stage and outcome compared with solitary symptoms¹⁷⁷. It is not possible to determine with certainty that the symptom recorded was due to cancer, although this is an issue with any study that describes the symptoms before diagnosis. This may explain the finding that delays with abdominal pain were associated with an improved outcome, as the initial presentation may have not been cancer related. The non-significant findings regarding stage

and delays for most of the non-specific patients may reflect a lack of power for these subsets. 20% of all patients had missing stage data, which previous studies have included as either stage III or IV. If the same was to be carried out in this study, it is likely to reinforce my findings as the referral interval III group for both colon and rectal cancer patients had the highest proportion of missing stage data compared with those referred earlier.

Analysis of patients without referral data suggests there is a reduced representation of colon and late stage cancers, as well as diagnoses in year before 2006. There are some possible reasons for these findings. The quality of both CPRD and the cancer registry has improved with time, and so patient diagnosed later have more complete records. Alternatively the medcodes used to identify referrals are less likely to identify the late stage cancer patients because late stage cancer is associated with non-specific symptoms (as shown in this chapter and again later). As mentioned previously non-specific symptoms are more likely than red-flag symptoms to be found in the free text which is not available for research. The referral codes were based on either the a Read code that specifically expresses the speciality or the presence of the symptom. The former should not produce a bias against late stage cancer. If there are patients being referred with non-specific symptoms, then these referrals could be missed.

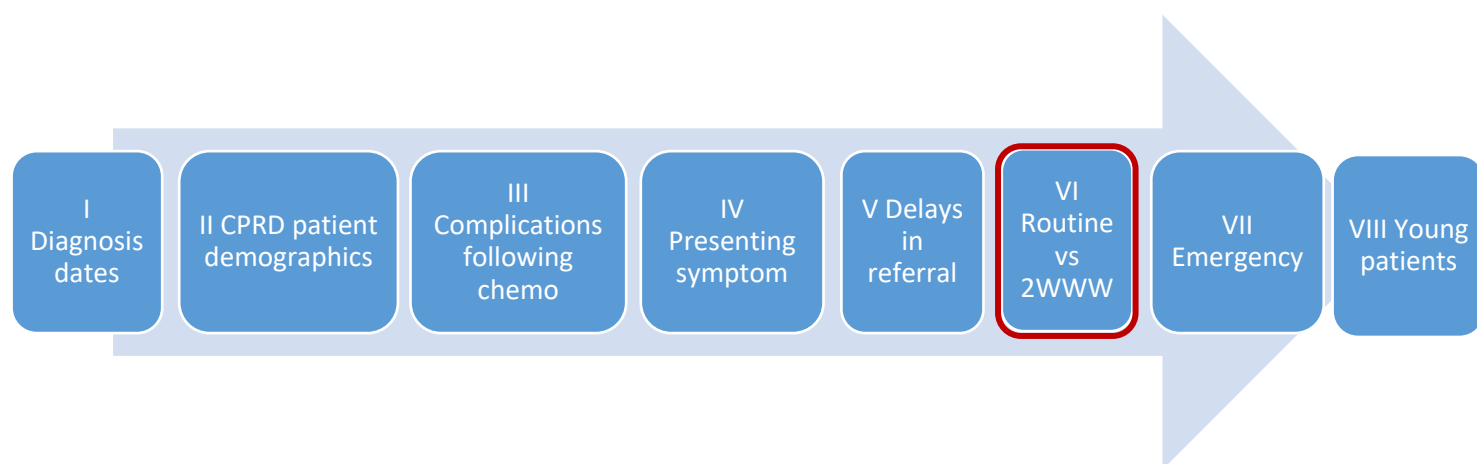
4.5.4.5 Implications for clinical practice

The importance of understanding the impact of delays in referral is underlined by an ongoing national study led by Cancer Research UK, which is collecting data from primary care regarding the number of consultations before referral and its association with outcome¹⁷⁸. Our study adds to this growing area of research. As rectal bleeding is an obvious deviation from normal health, and is a target to improve public awareness of bowel cancer, such patients seek medical help sooner than those with non-specific symptoms³⁶ leading to an earlier stage at diagnosis. Our study suggests delays in referral represent a missed opportunity to diagnose these patients. In contrast, the opportunity for an earlier diagnosis for patients with non-specific symptoms seems to be restricted as they are more likely to present at a later stage.

4.5.4.6 Conclusion

Patients with red-flag symptoms, and in particular rectal bleeding, are presenting at an earlier stage. The first consultation with the GP provides an opportunity to refer, that, if missed, could lead to a worse stage and prognosis for colorectal cancer patients.

End of Study



4.6 Study VI - Comparison of symptoms and prognosis between two-week-wait and non-urgent referrals

In the previous chapter I described the potentially worse prognosis for patient presenting with red-flag symptoms who wait over three months for a referral. In the following chapter I concentrate on whether there is a difference in outcome for patients who have been referred either through the two-week-wait or non-urgent pathways.

The hypothesis of this study states:

- **Patients referred via the two-week-wait pathway demonstrate a significantly better prognosis than patients referred via the non-urgent pathway**

I aim to answer the following research questions:

- **What is the difference in patient demographics and stage between the two pathways?**
- **Which symptoms are most likely to lead to a non-urgent referral?**
- **Does the hospital interval differ according to the referral pathway?**
- **After adjusting for patient demographics, the presenting symptom and the hospital interval, is there a difference in prognosis between the two pathways?**

4.6.1 Introduction

The Department of Health in 2000, followed by NICE in 2005, produced guidelines for an urgent referral for suspected colorectal cancer based on symptoms or anaemia. These were implemented to reduce the primary care interval and prevent emergency diagnoses for colorectal cancer patients with the ultimate intention to improve stage at diagnosis and prognosis. However the actual impact remains debatable. In this chapter I describe the differences in presenting symptoms, stage and prognosis between the two pathways, to determine whether the national referral guidelines have had the desired effect.

4.6.2 Method

4.6.2.1 Patient criteria

Only patients with a diagnosis in the cancer registry from 2006 onwards were selected as the route of diagnosis was available from this date. Patients under the age of 50 were excluded as the 2005 NICE referral criteria would not apply to the majority of these patients. The referral route was split into those identified as Two-week-wait (TWW) and non-urgent, the latter including 'GP referral' and 'other outpatient' routes only. Death certificate only, screening, emergency, unknown and inpatient elective were excluded as GP input before diagnosis would have been variable. In addition each patient should have had a referral in the year leading up to the diagnosis date and a relevant symptom in the year before referral.

4.6.2.2 Patient demographics and symptoms

The age, gender, IMD and Charlson score were compared between the referral routes, as was the stage and tumour site. The symptom at first consultation, and presence of the symptom before referral were also compared. This was carried out for symptoms individually and as a group of red-flag or non-specific. Significance was analysed with Chi-square analysis with testing of residuals to check for difference within categories.

4.6.2.3 Time to referral and diagnosis

The median time in days from presentation to referral (referral interval), from referral to diagnosis (hospital interval) and from presentation to diagnosis (diagnostic interval) were

compared between the routes using the Mann U Whitney test. The proportion of patients referred within two weeks, two weeks to three months and over three months was also compared using Chi square analysis.

Binary logistic modelling, using the Generalised Estimating Equation module, was used to determine the patient demographic most likely to lead to a two-week-wait referral while taking into account tumour site.

The above was repeated in a sub-group analyses for patients presenting with the two most common red-flag and non-specific symptoms, rectal bleeding or abdominal pain respectively.

4.6.2.4 Prognosis

Cox regression modelling provided the hazards ratio of death between the two-week-wait and non-urgent pathway while taking into account patient demographics and tumour site. The model was repeated to compare referral of over 3 months to those referred within 2 weeks, with inclusion of a continuous variable to account for the difference in hospital interval. A further model was carried out while only including patients referred within two weeks to reduce the impact of the referral interval.

A cox regression subgroup analysis was carried out including only patients presenting with rectal bleeding or abdominal pain.

4.6.3 Results

4.6.3.1 Study cohort

In total 3941 patients were identified to form the study cohort, of which 2165 (54.9%) and 1776 (45.1%) were diagnosed through the two-week-wait (TWW) and non-urgent pathways respectively. There was no significant difference between the routes in terms of age, gender or IMD (table 40). A higher proportion of patients with a Charlson score of one (45.7% vs 54.3%) or more than one (45.7% vs 54.3%) were diagnosed through the non-urgent pathway. The majority of both colon and rectal cancer patients were diagnosed through the

2WW pathway, although rectal cancer patients demonstrated a higher percentage (59.3% vs 52.4%). There was no difference in the proportion of stage I disease between the two routes. A significantly higher proportion of cancers diagnosed through the TWW were of stage II (28.5% vs 25.4%), III (28.9% vs 25.6%) or IV (14.8% vs 12.5%), but lower missing stage (14.4% vs 21.2%).

Table 40 - Patient demographics and tumour characteristics according to the route of diagnosis

		All		Two week wait		Non-urgent		p
		n	% of N	n	% of N	n	% of N	
Age	50 - 59	549	13.9	289 _a	52.6	260 _a	47.4	0.598
	60 - 69	982	24.9	544 _a	55.4	438 _a	44.6	
	70 - 79	1389	35.2	759 _a	54.6	630 _a	45.4	
	80 and over	1021	25.9	573 _a	56.1	448 _a	43.9	
Gender	Male	2182	55.4	1212 _a	55.5	970 _a	44.5	0.391
	Female	1759	44.6	953 _a	54.2	806 _a	45.8	
IMD	1	984	25.0	532 _a	54.1	452 _a	45.9	0.538
	2	1038	26.3	565 _a	54.4	473 _a	45.6	
	3	797	20.2	458 _a	57.5	339 _a	42.5	
	4	682	17.3	365 _a	53.5	317 _a	46.5	
	5	440	11.2	245 _a	55.7	195 _a	44.3	
Charlson score	Zero	3036	77.0	1732 _a	57	1304 _b	43	< 0.0005
	One	570	14.5	280 _a	49.1	290 _b	50.9	
	Two or more	335	8.5	153 _a	45.7	182 _b	54.3	
Site	Colon	2517	63.9	1320 _b	52.4	1197 _a	47.6	< 0.0005
	Rectal	1424	36.1	845 _b	59.3	579 _a	40.7	
Stage	I	566	14.4	291 _a	13.4	275 _a	15.5	< 0.0005
	II	1068	27.1	617 _a	28.5	451 _b	25.4	
	III	1079	27.4	625 _a	28.9	454 _b	25.6	
	IV	543	13.8	321 _a	14.8	222 _b	12.5	
	Not provided	685	17.4	311 _a	14.4	374 _b	21.1	

Subscripts represent differences in category

4.6.3.2 Presenting symptom and symptom before referral

Of the 1551 patients presenting with red-flag symptoms, 58.1% were referred through the 2WW pathway (table 41). Of the patients with a red-flag before referral, 57.6% were referred through this route. Presence of an abdominal mass at presentation or before referral was associated with the highest proportion of 2WW referrals (69% and 72.1% respectively). The presence of rectal bleeding at first presentation (55.6% vs 44.4%) or before referral (54% vs 46%) did not significantly alter the percentage of referrals via the

TWW. This was similar to anaemia and diarrhoea. However a significantly higher proportion of patients with a change in bowel habit at first consultation or before referral were diagnosed through a TWW pathway (66.9% and 66.3% respectively). Of the non-specific symptoms, presenting with abdominal pain, other bowel function and weight loss or fatigue was associated with a higher proportion of non-urgent referrals (52.4%, 59.5%, 51.7% respectively). There was no significant difference if constipation was the presenting symptom. However a significantly higher proportion of patients with constipation at any time before referral were referred through the non-urgent pathway (52.1 vs 47.9% $p = 0.004$).

Table 41 The referral pathway according to presenting symptom and symptom before referral

	At first presentation					Before referral				
	Two week wait		Non-urgent		p	Two week wait		Non-urgent		p
	n	% of N	n	% of N		n	% of N	n	% of N	
Red-flag	1551	58.1	1121	41.9	< 0.0005	1785	57.6	1316	42.4	< 0.0005
Rectal bleeding	590	55.6	472	44.4	0.635	694	54	592	46	0.395
Anaemia	331	54.3	279	45.7	0.132	442	54.6	368	45.4	0.814
Change in bowel habit	351	66.9	174	33.1	< 0.0005	490	66.3	249	33.7	< 0.0005
Diarrhoea	263	58.4	187	41.6	0.112	346	58.3	247	41.7	0.07
Abdominal mass	20	69	9	31	0.127	49	72.1	19	27.9	0.004
Abdominal pain	314	47.6	345	52.4	< 0.0005	391	47.1	440	52.9	< 0.0005
Constipation	147	51.6	138	48.4	0.237	182	47.9	198	52.1	0.004
Other bowel function	32	40.5	47	59.5	0.009	51	42.1	70	57.9	0.004
Weight loss/fatigue	117	48.3	125	51.7	0.033	178	50.7	173	49.3	0.096

4.6.3.3 Time to referral and diagnosis

Patients diagnosed via the non-urgent route demonstrated a significantly longer referral (median 18 IQR 1 – 111), hospital (68 IQR 37 – 112) and diagnostic (116 IQR 62 – 235) interval than those referred via the TWW (table 42). A significantly lower proportion of non-urgent patients were referred within 2 weeks (47.6% vs 59.6%).

Table 42 – Referral, hospital and diagnostic interval (median with IQR) and the percentage of patients referred within 2 weeks, 2 weeks to 3 months and over 3 months according to diagnostic route

		2WW		Non-urgent		
Interval	Referral	6 (0 – 54)		18 (1 – 111)		< 0.0005
	Hospital	24 (14 – 40)		68 (37 – 112)		< 0.0005
	Diagnostic	43 (22 – 99)		116 (62 – 235)		< 0.0005
Referral time	Within 2 weeks	1290 ^a	59.6	846 ^b	47.6	< 0.0005
	2 weeks – 3 months	454 ^a	21.0	432 ^b	24.3	
	Over 3 months	421 ^a	19.4	498 ^b	28.0	

4.6.3.4 Regression analysis to determine the odds ratio for a TWW referral

Patients aged between 50 – 59 were 16% less likely to be referred through the 2WW pathway compared to 60 – 69 year olds (OR 0.84 95% CI 0.65 – 0.98 p = 0.034). No other age group, IMD or gender demonstrated a significantly different odds ratio. Patients with colon cancer (OR 0.77 95% CI 0.67 – 0.88 p < 0.0005), Charlson score of more than one (OR 0.62 95% CI 0.50 – 0.76 p < 0.0005) and score of one (OR 0.72 95% CI 0.59 – 0.87 p = 0.001) were less likely to be referred through the TWW route.

The presence of a red-flag symptom at the first consultation increased the odds ratio of a TWW referral by 42% (OR 1.42 95% CI 1.24 – 1.64 p < 0.0005) (table 43). The presence of a red-flag symptom at any time before the decision to refer increased the odds ratio to 1.56 (95% CI 1.32 – 1.85 p < 0.0005). Compared to patients presenting with rectal bleeding, only a change in bowel habit was associated with increased odds ratio of a TWW referral (OR 1.60 95% CI 1.29 – 2.00). On the contrary presenting with abdominal pain (OR 0.79 95% CI 0.65 – 0.97 p = 0.021) or other bowel function (OR 0.58 95% CI 0.37 – 0.92 p = 0.019) reduced the likelihood of a 2WW referral. The presence of rectal bleeding at any time before referral reduced the odds of a TWW (OR 0.87 95% CI 0.76 – 0.99 p = 0.031), as did abdominal pain, constipation or ‘other bowel function’.

Table 43 – Odds ratio (OR) of a two week wait referral based on initial symptom (rectal bleeding as reference) and presence of symptom before referral (absence of symptom as reference)

	At first presentation		At any time	
	OR	p	OR	p
Red flag	1.42 (1.24 – 1.64)	< 0.0005	1.56 (1.32 – 1.85)	< 0.0005
Rectal bleeding	ref		0.87 (0.76 – 0.99)	0.031
Anaemia	1.06 (0.86 – 1.31)	0.596	1.11 (0.94 – 1.31)	0.234
Change in bowel habit	1.60 (1.29 – 2.00)	< 0.0005	1.70 (1.44 – 2.02)	< 0.0005
Diarrhoea	1.14 (0.91 – 1.43)	0.256	1.15 (0.95 – 1.39)	0.147
Mass	1.84 (0.84 – 4.05)	0.127	2.15 (1.28 – 3.63)	0.004
Abdominal pain	0.79 (0.65 – 0.97)	0.021	0.71 (0.60 – 0.84)	< 0.0005
Constipation	0.86 (0.66 – 1.11)	0.250	0.72 (0.59 – 0.88)	0.001
Other bowel function	0.58 (0.37 – 0.92)	0.019	0.62 (0.43 – 0.88)	0.008
Weight loss or fatigue	0.82 (0.61 – 1.10)	0.181	0.88 (0.70 – 1.10)	0.248

Both models include age, gender, cancer site, IMD and Charlson score as co-variables.

4.6.3.5 Prognosis between TWW and non-urgent referral patients

After accounting for age, gender, IMD, Charlson score and site of cancer, there was no significant difference in patients referred by the 2WW route (HR 0.91 95% CI 0.81 – 1.01 p = 0.074). After taking into account the presenting symptom, once again there was no difference seen between the two routes (HR 0.92 95% CI 0.83 – 1.03 p = 0.136). In a third model after accounting for stage, the non-significant difference was maintained (HR 0.91 95% CI 0.82 – 1.01 p = 0.91). In all three models, age of over 80 was associated with a significantly worse prognosis than those aged 60 – 69, as was increasing Charlson score and higher IMD. Gender had no impact on prognosis. After including only patients that were referred within 2 weeks, but accounting for the above variables, again no difference was seen in prognosis between the TWW and non-urgent pathways.

4.6.3.6 Subgroup analysis – Presenting with rectal bleeding, abdominal pain

4.6.3.6.1 Rectal bleeding

For all patients presenting with rectal bleeding, there was no significant difference in age, gender, Charlson score or IMD in univariate analysis (table 44). Patients referred via the TWW had a significantly higher stage III but lower stage ‘unknown’ compared to the non-urgent route. In multi-variate analysis, only patients with Charlson score of two or more were more likely than those scoring zero to be referred via the non-urgent route than the two week wait route (table 45).

Table 44 - Demographics and tumour stage according to route of diagnosis for patients with rectal bleeding at first presentation

		Two week wait		Non-urgent		p	Total	
		n	%	n	%		N	%
Age	50 - 59	113 _a	55.7	90 _a	44.3	0.898	203	19.1
	60 - 69	164 _a	53.8	141 _a	46.2		305	28.7
	70 - 79	201 _a	56.5	155 _a	43.5		356	33.5
	80 and over	112 _a	56.6	86 _a	43.4		198	18.6
Gender	Male	348 _a	55.9	274 _a	44.1	0.759	622	58.6
	Female	242 _a	55.0	198 _a	45.0		440	41.4
Charlson score	Zero	495 _a	56.7	378 _a	43.3	0.095	873	82.2
	One	54 _a	46.2	63 _b	53.8		117	11.0
	Two or more	41 _a	56.9	31 _a	43.1		72	6.8
Site of cancer	Colon	261 _a	51.8	243 _b	48.2	0.019	504	47.5
	Rectum	329 _a	59.0	229 _b	41.0		558	52.5
IMD	1	155 _a	54.6	129 _a	45.4	0.398	284	26.7
	2	156 _a	54.0	133 _a	46.0		289	27.2
	3	111 _a	55.0	91 _a	45.0		202	19.0
	4	100 _a	54.9	82 _a	45.1		182	17.1
	5	68 _a	64.8	37 _b	35.2		105	9.9
Stage	I	117 _a	19.8	114 _a	24.2	0.004	231	21.8
	II	145 _a	24.6	96 _a	20.3		241	22.7
	III	183 _a	31.0	119 _b	25.2		302	28.4
	IV	60 _a	10.2	42 _a	8.9		102	9.6
	Not provided	85 _a	14.4	101 _b	21.4		186	17.5

Subscripts represent significant differences between TWW and non-urgent groups for that variable

Table 45 - Odds ratio of a two week wait referral for patients presenting with rectal bleeding

		OR	p
Age	50 - 59	1.02 (0.71 - 1.46)	0.905
	60 - 69	ref	
	70 - 79	1.13 (0.81 - 1.56)	0.480
	80 and over	1.19 (0.83 - 1.73)	0.346
Gender	Male	1.04 (0.81 - 1.35)	0.752
	Female	ref	
Charlson score	Zero	ref	
	One	0.97 (0.59 - 1.61)	0.908
	Two or more	0.64 (0.44 - 0.94)	0.023
Site of cancer	Colon	0.76 (0.60 - 0.96)	0.02
	Rectum	ref	
IMD	1	ref	
	2	1.016 (0.74 - 1.40)	0.92
	3	1.03 (0.70 - 1.56)	0.87
	4	1.05 (0.72 - 1.54)	0.79
	5	1.54 (0.97 - 2.45)	0.07

Although there was a significant difference in the median time to referral between the two routes, the difference was only a median of 2 days (table 46). However the hospital interval was considerably longer, with a median of 63.5 days for the non-urgent route and 21 days for the TWW route. This was reflected in the longer diagnostic interval.

Table 46 – Primary care, hospital and total diagnostic interval for TWW and non-urgent patients presenting with rectal bleeding

	Two week wait	Non-urgent	p
Primary care interval	0 (0 – 8)	2 (0 – 15)	< 0.05
Hospital interval	21 (12.75 – 33)	63.5 (37 – 103)	< 0.05
Diagnostic interval	27 (14 – 60)	85.5 (47 – 140.75)	< 0.05

Cox regression analysis demonstrated no difference in outcome between TWW and non-urgent routes for patients presenting with rectal bleeding (table 47). However, as previously noted, patients referred after 3 months had a significantly worse outcome than those referred within 2 weeks.

Table 47 – Cox regression modelling for survival following referral via two week wait referral, and for all patients dependent on referral delay

	Model 1		Model 2		Model 3	
	HR	p	HR	p	HR	p
Two week wait	1.04 (0.82 – 1.32)	0.738	0.81 (0.64 – 1.13)	0.09	0.76 (0.58 – 1.03)	0.056
Urgent	ref		ref			
Within 2 weeks	-	-	ref		ref	
2 weeks – 3 months	-	-	1.23 (0.86 – 1.76)	0.252	1.23 (0.86 – 1.75)	0.263
Over 3 months	-	-	1.94 (1.40 – 2.70)	< 0.0005	1.97 (1.42 – 2.74)	< 0.0005

Model 1 includes age, gender, tumour site, Charlson score and IMD). Model 2 = Model 1 + time to referral (categorical),

Model 3 = Model 2 + time from referral to diagnosis (continuous)

4.6.3.6.2 Abdominal pain

For all patients presenting with abdominal pain, there was no significant difference in age, gender, Charlson score or IMD in univariate analysis (table 48). A significantly lower proportion of patients referred via the TWW were diagnosed with stage I disease compared with the non-urgent route (5.1% vs 9.3%). However there was no overall difference in the distribution of stage between the two routes. In multi-variate analysis, only patient with Charlson score of two or more were more likely than those of Charlson score of zero to be referred via the non-urgent route than the TWW route (table 49).

Table 48 – Patient demographics and stage of cancer according to referral route for patients presenting with abdominal pain

		Two week wait		Non-urgent		p	Total	
		n	%	n	%		n	%
Age	50 - 59	48 _a	42.1	66 _a	57.9	0.617	114	17.3
	60 - 69	99 _a	49.3	102 _a	50.7		201	30.5
	70 - 79	111 _a	48.1	120 _a	51.9		231	35.1
	80 and over	56 _a	49.6	57 _a	50.4		113	17.1
Gender	Male	166 _a	49.8	167 _a	50.2	0.253	333	50.5
	Female	148 _a	45.4	178 _a	54.6		326	49.5
Charlson score	Zero	251 _a	48.9	262 _a	51.1	0.103	513	77.8
	One	46 _a	48.4	49 _a	51.6		95	14.4
	Two or more	17 _a	33.3	34 _b	66.7		51	7.7
Site of cancer	Colon	257 _a	46.1	301 _a	53.9	0.055	558	84.7
	Rectum	57 _a	56.4	44 _a	43.6		101	15.3
IMD	1	70 _a	45.8	83 _a	54.2	0.384	153	23.2
	2	93 _a	50.8	90 _a	49.2		183	27.8
	3	75 _a	51.4	71 _a	48.6		146	22.2
	4	43 _a	40.2	64 _a	59.8		107	16.2
	5	33 _a	47.1	37 _a	52.9		70	10.6
Stage	I	16 _a	5.1	32 _b	9.3	0.110	48	7.3
	II	95 _a	30.3	100 _a	29.0		195	29.6
	III	94 _a	29.9	93 _a	27.0		187	28.4
	IV	66 _a	21.0	59 _a	17.1		125	19.0
	Not provided	43 _a	13.7	61 _a	17.7		104	15.8

Subscripts represent differences between groups at $p < 0.05$

Table 49 – Odds ratio of a Two Week Wait referral for patients presenting with abdominal pain

		OR (95% CI)	p
Age	50 - 59	0.71 (0.44 - 1.13)	0.149
	60 - 69	ref	
	70 - 79	0.99 (0.67 - 1.46)	0.958
	80 and over	1.01 (0.62 - 1.66)	0.958
Gender	Male	1.25 (0.89 – 1.75)	0.195
	Female	ref	
Charlson score	Zero	ref	
	One	0.95 (0.61 – 1.49)	0.823
	Two or more	0.52 (0.29 – 0.93)	0.028
Site of cancer	Colon	0.64 (0.40 – 1.03)	0.065
	Rectum	ref	
IMD	1	ref	
	2	1.18 (0.75 - 1.87)	0.470
	3	1.23 (0.78 - 1.94)	0.384
	4	0.79 (0.47 - 1.33)	0.373
	5	1.08 (0.60 – 1.49)	0.823

There was a significant difference in the median time to referral between the two routes of a median of 14.5 days (table 50). The hospital interval was also considerably longer, with a median of 70 days for the non-urgent route and 28 days for the TWW route. This was reflected in the longer diagnostic interval.

Table 50 – Primary care, hospital and total diagnostic interval for patients presenting with abdominal pain

	Two week wait	Non-urgent	p
Primary care interval	27.5 (2.75 – 126.25)	42 (9 – 168)	< 0.05
Hospital interval	28 (16 – 49)	70 (40 – 113.5)	< 0.05
Diagnostic interval	68.5 (37 – 182.25)	147 (78.5 – 270)	< 0.05

Cox regression analysis demonstrated no difference in outcome between TWW and non-urgent route patients (table 51). However, as previously noted, patients referred after 3 months had a significantly better outcome than those referred within 2 weeks.

Table 51 - Cox regression modelling for survival following referral via two week wait referral, and for all patients dependent on referral delay, for patients presenting with abdominal pain

	Model 1		Model 2		Model 3	
	HR	p	HR	p	HR	p
Two week wait	0.79 (0.62 – 1.01)	0.059	1.02 (0.80 – 1.29)	0.896	0.97 (0.75 – 1.26)	0.820
Urgent	ref		ref		ref	
Within 2 weeks	-	-	ref		ref	
2 weeks – 3 months	-	-	0.98 (0.74 – 1.29)	0.875	0.98 (0.75 – 1.29)	0.897
Over 3 months	-	-	0.59 (0.43 – 0.80)	0.001	0.58 (0.43 – 0.79)	0.001

Model 1 = age, gender, tumour site, Charlson score and IMD. Model 2 = Model 1 + time to referral (categorical), Model 3 = Model 2 + time from referral to diagnosis (continuous)

After accounting for age, Charlson score, gender and IMD, the HR of cancer related mortality for patients with a red-flag at first presentation, referred via the non-urgent pathway who had experienced a delay in referral of over 3 months compared to those referred within 2 weeks, was HR 1.32 (95% CI 1.02 – 1.70 $p = 0.036$). For patients referred by the 2WW pathway, also presenting with a red-flag symptoms, the HR was 1.41 (95% CI 1.11 – 1.80 $p = 0.005$).

4.6.4 Discussion

4.6.4.1 Summary

Through the analysis of over 3000 colorectal cancer patients aged over 50, this study demonstrates the symptoms most associated with a TWW rather than a non-urgent referral and the stage and prognosis difference between the two groups. The key determinant of the route of referral was the symptom at first presentation or leading up to referral, although patients aged 50 – 59 were less likely to be referred urgently. Of note, neither route demonstrated a better prognosis, even though the hospital and total diagnostic route was considerably longer for the non-urgent patients. This same pattern was repeated when only including patients with the same presenting symptom.

4.6.4.2 Relevance to clinical practice

The NICE referral guidelines were introduced to aid GPs to identify cancer patients earlier by describing a set of symptoms that if present, should trigger an urgent pathway⁵². This policy was implemented in response to the relatively poor outcome noted in the UK compared to the Northern European countries. On receipt of this urgent referral, the onus is on secondary care to provide a clinic consultation within two weeks⁴⁹.

This study demonstrates 54.9% of patients were referred through the TWW pathway. The symptoms with the highest proportion of TWW referrals, closely follows the 2005 NICE guidelines - 55.6% of patients presenting with rectal bleeding and 66.9% to those with a change in bowel habit were referred through the TWW, which reduced to 47.6% for abdominal pain and 40.5% for 'other bowel function'. Although the majority of rectal bleeding patients were referred as TWW (55.6%), there is a considerably high percentage who, despite attending with a red-flag symptoms, were still referred via the non-urgent routes. There are two possible reasons for this. Either patients were presenting with symptom combinations or durations which failed to fulfil the referral criteria, or the GPs were not aware of the referral criteria. Considering the effort applied to make the 2005 guidelines accessible to all, it is unlikely the latter issue is a significant factor. Unfortunately the free text from CPRD is not available for research purposes, so it is not possible to determine the duration of symptoms. Patient demographics do not seem to be a determining factor, except those aged between 50 – 59. With an age threshold of 60 years age for some of the referral criteria, it is understandable that those aged 50 – 59 were less likely than those aged 60 – 69 to undergo an urgent referral.

Survival analysis demonstrates the route of referral is not a major factor on prognosis or stage, which negates the hypothesis that a TWW referral is better for the patient. This finding was seen consistently even after only including patients with the same presenting symptom, in order to exclude the effect of the type of symptom on survival. The time to refer significantly altered prognosis only if it was delayed by over three months, and only for those presenting with red-flag symptoms such as rectal bleeding.

Compared with the NCIN data⁹⁹, the CPRD linked data presented here also showed TWW patients had a lower proportion of Stage I disease, but higher Stage III and less patients had missing stage data. This is the largest study to consider the impact of symptoms, not only on survival, but the likelihood of a TWW referral. As with previous studies we found no association between the route and survival. A greater impact seems to be the length of time to referral, and the symptom at presentation.

Previously a waiting time paradox has been described, where patients who present with short history with severe symptoms are referred urgently, as opposed to those with a long history of indolent symptoms⁸⁰. The former is thought to be due to aggressive tumours, and would explain the relatively lower stage I disease in TWW referral for all patients and those with rectal bleeding. This is balanced by the higher stage 'unknown', which has previously been considered late stage disease due to the relatively poor outcome noted for these cases in epidemiological slides⁹⁷.

4.6.4.3 Limitations

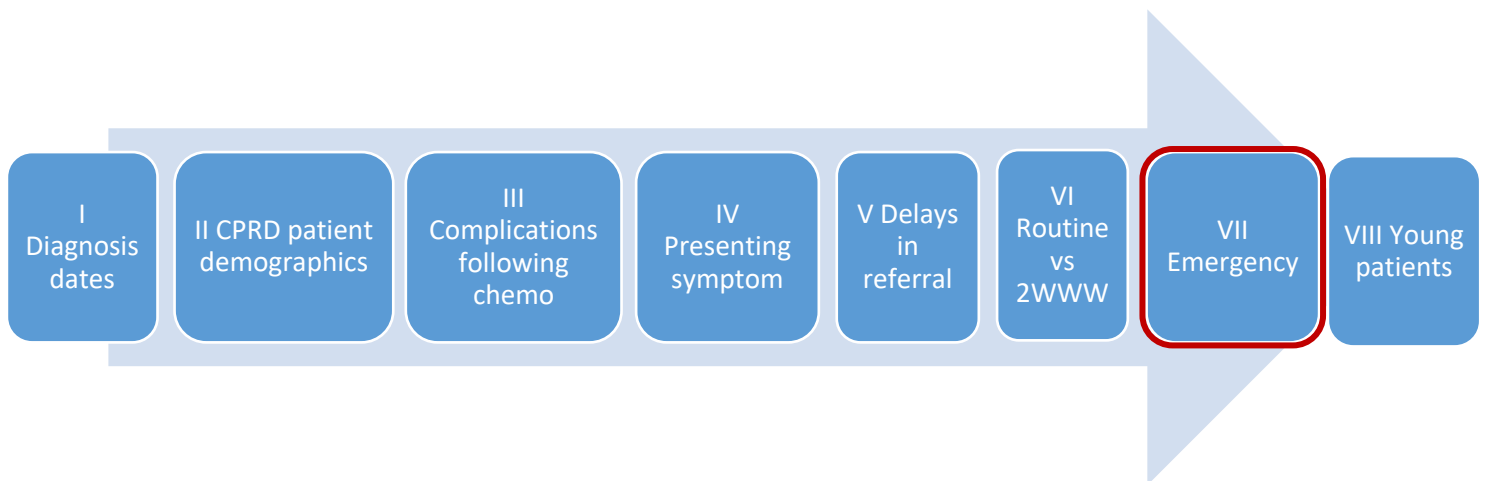
Patients were selected for this study based on the route of diagnosis, and so were all expected to have seen their GP and have been referred. As stated previously it is difficult to determine whether the recorded symptom is due to an underlying cancer. Similarly it is not possible to conclude with certainty that the referral considered was the referral that led to the diagnosis. It is possible that the patient was referred non-urgently for a non-related condition, and re-referred as an urgent due to red-flag symptoms. Such patients would be considered as non-urgent patients with a long hospital interval. The proportion of patients falling into this category would be expected to be low as referral medcodes were chosen based on the likelihood of cancer related.

4.6.4.4 Implications for health policy

The benefit of the referral guidelines has been to highlight certain symptoms that are more likely cancer related, rather than the route of referral. The criteria are based on the

likelihood of cancer, rather than the prognosis following diagnosis^{49,50}. Without the referral guidelines, it is possible more patients would wait a considerable amount of time before referral, potentially reaching the detrimental three month threshold. The most recent guidelines of 2015 have reduced the PPV by excluding the need for the duration of symptoms⁵⁰. This is likely to increase the number of patients being referred, and increase the burden of the TWW pathway without any benefit to the patient. On the contrary the shift of focus and resources to the increased TWW pathway could further prolong the diagnostic interval for non-urgent referrals, thereby worsening their outcome. Patients would understandably want to be referred and seen in secondary care urgently. However this study adds to previous studies that there is insufficient evidence to suggest the arbitrary two week criteria improves outcome for the patient^{66,71}.

End of Study



4.7 Study VII - Missed opportunities in primary care to prevent an emergency diagnosis

In the previous chapter I described the impact of the urgent pathway on the outcome of colorectal cancer patients. In this study I switch the focus to emergency patients, and determine what led to missed opportunities for a non-emergency diagnosis.

The hypothesis states:

- **Patients experience missed opportunities in primary care to prevent an emergency diagnosis**

The research questions are:

- **Which patients are most likely to be diagnosed as an emergency?**
- **Which symptoms are most likely to lead to an emergency diagnosis after accounting for patient demographics?**
- **Is there a difference in GP usage between emergency and non-emergency patients in the four years leading up to diagnosis?**
- **Could emergency patients have been identified earlier to prevent an emergency diagnosis?**
- **Are emergency patients with the same symptoms as non-emergency patients diagnosed at a similar tumour stage?**

4.7.1 Introduction

Despite the introduction of the national referral guidelines, the percentage of colorectal cancer patients diagnosed as an emergency has not changed tremendously⁹⁸. Factors associated with an emergency diagnosis can be considered non-modifiable, such as aggressive tumours without previous symptoms to herald the diagnosis, and potentially modifiable factors such as missed opportunities in primary to identify the cancer⁸⁸⁹³. This study focuses on the latter. By using CPRD, it is possible to identify differences in presenting symptoms between non-emergency and emergency cancer patients after adjusting for patient demographics. A better understanding around missed opportunities in primary care has the potential to improve prognosis.

4.7.2 Method

4.7.2.1 Patient cohort

Patients were identified as having a colorectal cancer diagnosis between 2006 and 2013 inclusive, as the route of diagnosis was available for these years. Patients diagnosed through death certificate only, unknown route or screening or those that had a previous cancer diagnosis in any one of the three linked datasets were excluded. Colon and rectal cancers were analysed separately as preliminary analysis suggested these present with different symptoms, stage and age.

4.7.2.2 Definition of diagnostic routes

The emergency diagnosis route included patients referred as an emergency by the GP. Therefore the last 28 days before a colorectal cancer diagnosis were not considered a time frame during which a missed opportunity for diagnosis would have occurred. Instead these consultations were assumed to have led directly to the diagnosis. This criteria has been previously described¹⁰⁵. The non-emergency route was an amalgamation of 'two week wait', 'other outpatient', 'GP referral' and 'inpatient elective'.

4.7.2.3 Statistical analysis

Univariate analysis was used to compare all non-emergency and emergency patients in terms of age, gender, IMD, Charlson score and stage. The proportion of non-emergency diagnoses according to the individual routes is described.

4.7.2.4 Number of consultations before diagnosis

The number of consultations with relevant symptoms and consultations for any reason within the penultimate 28 days, between 28 days to a year, one year to two years, two to three years and three to four years was compared between emergency and non-emergency routes with univariate analysis. Relevant consultations were those with a symptom of interest recorded. Duplicate consultations based on consultation ID and date were excluded for the purpose of counting, but were used to determine the presence of symptoms. Therefore a patient may have had more than one symptom recorded.

The consultation rate per week, defined as the average number of consultations per patient per week, were compared in the year leading up to the colorectal cancer diagnosis (minus the last 28 days) between emergency and non-emergency routes. This was carried out for all consultations and repeated only for consultations with relevant symptoms.

4.7.2.5 Patient factors and symptoms associated with an emergency diagnosis

Binary logistic regression models were constructed, using the Generalised Estimating Equation module, with emergency diagnosis as the dependent factor, for the following:

Model I: determine the age, gender, IMD and Charlson score favouring an emergency diagnosis. The model accounted for the number of consultations for all reasons between 28 days and one year as a continuous variable.

Model II: determine which symptoms were associated with an emergency diagnosis after accounting for the variables in model I. Symptoms were included individually as a binary

categorical variable (present or absent). This model was repeated with symptoms as a binary categorical variable as red-flag and non-specific only.

A binary logistic regression model, again using the Generalised Estimating Equation module, was constructed with stage as the dependent factor, for the following:

Model III: determine whether emergency and non-emergency patients presenting with the same symptoms demonstrate a different stage at diagnosis. Late stage was defined as stage III, IV or missing. Six models were constructed including only patients with the following symptoms: rectal bleeding, anaemia, abdominal pain, constipation, red-flag, non-specific. This was repeated for colon and rectal separately. Previous studies have included missing stage as late⁹⁷ as these patients have been shown to have a similar worse prognosis⁹².

4.7.2.6 Comparison of patients with missed opportunity for non-emergency diagnosis with those with an emergency

Patients who experienced a missed opportunity for a non-emergency diagnosis were defined as the study cohort, with the criteria of:

1. diagnosed through the emergency route and
2. relevant symptom recorded in between 28 days to one year before diagnosis

The control cohort criteria were:

1. diagnosed through the non - emergency route and
2. relevant symptom recorded in between 28 days to one year before diagnosis

Univariate analysis was used to compare the study and control cohort in terms of age, gender, IMD, Charlson score, stage and symptom between 28 days to one year before diagnosis.

Binary logistic regression models I and II were constructed as per study part one but for these two cohorts.

4.7.3 Results

4.7.3.1 Route to diagnosis

Between 2006 and 2013 inclusive, there were 10692 colorectal cancer diagnosis, of which 7671 (71.7%) were colon, and 3021 (28.3%) were rectal. Of these, 2417 (31.5%) and 373 (12.3%) were diagnosed as an emergency respectively (table 52). Of the non-emergency routes, the TWW was the most common for both colon (29.7%) and rectum (45.1%) cancers, followed by the 'GP referral' route (table 27).

Table 52 – Route to diagnosis for non-emergency colorectal cancer cases between 2006 -13

	Colon		Rectum		All	
	n	%	n	%	n	%
TWW	2278	29.7	1363	45.1	3641	34.1
GP referral	1887	24.6	915	30.3	2802	26.2
Other outpatient	698	9.1	198	6.6	896	8.4
Inpatient elective	391	5.1	172	5.7	863	8.1
Emergency	2417	31.5	373	12.3	2820	26.4

TWW – Two week wait; GP = General Practitioner

4.7.3.2 Comparison of demographics and stage between emergency and non-emergency routes

A significantly higher proportion of female patients were diagnosed as an emergency for both colon (31.1% vs 28.6% $p < 0.005$) and rectal cancer (13.4% vs 10.9% $p = 0.032$) (table 53). Patients aged over the age of 80 demonstrated the highest proportion of emergency diagnoses (38.9% colon, 19.6% rectal). For those under the age of 50 patients, 34.9% of colon and 8.4% of the rectal cancer patients were diagnosed as an emergency. The lowest proportion of emergency diagnoses was seen for 60 – 69 year olds for colon cancer (25.2%) and 50 – 59 year olds for rectal (8.5%). Lower levels of socioeconomic deprivation were consistently associated with a higher proportion of emergency diagnoses for both cancers, as was higher Charlson scores. Emergency diagnosis was also associated with a higher stage. For colon cancer, 22.7% of emergency compared to 16.1% of non-emergency were stage IV, while 2.6% of emergency and 10.8% of non-emergency were stage I. For rectal cancer 23.5% of emergency compared to 12.8% of non-emergency were stage IV, while 8.2% of emergency and 17.3% of non-emergency were stage I. Emergency diagnoses were also more likely to have missing stage.

Table 53 - Proportion of patients with an emergency diagnosis according to gender, age, IMD and Charlson score between 2006 – 13 inclusive

		Colon					Rectum				
		Non-emergency		Emergency			Non-emergency		Emergency		
		n	%	n	%	p value	n	%	n	%	p value
Total		5662	69.7	2459	30.2		2821	88.2	379	11.8	
Gender	Male	2998	71.4	1199	28.6	< 0.005	1754	89.1	214	10.9	0.032
	Female	2664	67.9	1260	32.1		1067	86.6	165	13.4	
Age	Under 50	261	65.1	140	34.9	< 0.0005	174	91.6	16	8.4	< 0.0005
	50 - 59	625	74.7	212	25.3		452	94.0	29	6.0	
	60 - 69	1307	74.8	441	25.2		691	91.5	64	8.5	
	70 - 79	1903	74	668	26		871	88.2	116	11.8	
	80 and over	1566	61.1	998	38.9		633	80.4	14	19.6	
IMD	1	1403	72.8	524	27.2	< 0.0005	675	91.7	61	8.3	< 0.0005
	2	1480	72.5	560	27.5		703	89.7	81	10.3	
	3	1142	69.0	514	31.0		581	87.9	80	12.1	
	4	982	68.1	461	31.9		512	85.3	88	14.7	
	5	655	62.1	400	37.9		350	83.5	69	16.5	
Charlson	Zero	4187	74.9	1403	25.1	< 0.0005	2296	91.8	204	8.2	< 0.0005
	One	888	61.3	561	38.7		322	76.5	99	23.5	
	≥2	587	54.3	495	45.7		203	72.8	76	27.2	
Stage	I	614	10.8	63	2.6	< 0.0005	489	17.3	31	8.2	< 0.0005
	II	1706	30.1	582	23.7		509	18.0	35	9.2	
	III	1416	25.0	626	25.5		780	27.6	75	19.8	
	IV	910	16.1	557	22.7		361	12.8	89	23.5	
	Missing	1016	17.9	631	25.7		682	24.2	149	39.3	

4.7.3.3 Any consultations before diagnosis

A significantly higher proportion of colon cancer patients with an emergency diagnosis had a recording in CPRD of a consultation in the 28 days before diagnosis (98.1% vs 96.9%) (table 54). However between 28 days to one year, one to two years and two to three years, non-emergency patients demonstrated a significantly higher proportion of consultations. Beyond three years there was no difference. For rectal cancer, no difference was seen, irrespective of the time period before diagnosis.

Table 54 - Presence of a consultation for any reason according to time from diagnosis

		Colon			Rectum		
		Elective	Emergency	p value	Elective	Emergency	p value
Consultation within 28 days	n	5485	2413	< 0.005	2724	370	0.277
	%	96.9	98.1		96.6	97.6	
Between 28 days and one year	n	5593	2388	< 0.0005	2745	364	0.165
	%	98.8	97.1		97.3	96.0	
Between one and two years	n	5462	2338	< 0.005	2262	354	0.451
	%	96.5	95.1		94.3	93.4	
Between two and three years	n	5373	2299	0.011	2626	349	0.473
	%	94.9	93.5		93.1	92.1	
Between three and four years	n	5214	2242	0.168	2546	338	0.512
	%	92.1	91.2		90.3	89.2	

4.7.3.4 Consultations with relevant symptoms before diagnosis

A higher proportion of both colon (30.6%) and rectal cancer (38.8%) patients diagnosed as an emergency had no relevant consultations recorded in CPRD in the year before diagnosis (table 55) compared with non-emergency diagnoses (23.3% of colon and 21.8% of rectal). Colon cancer patients diagnosed as an emergency were significantly more likely to present to their GP in the 28 days before diagnosis (47.9% vs 28.7%) compared with non-emergency patients, but significantly less likely to present beyond this period up to a year (43.3% vs 61.9%). A higher proportion of colon cancer emergency patients were seen 3 to 4 times (5.4% vs 1.1%) and 5 or more time (0.5% vs 0.02%) within 28 days before diagnosis. For rectal cancer there was no significant difference in the 28 days leading up to diagnosis. However again the emergency patients were less likely to see their GP (39.1% vs 51.9%) between 28 days to one year. Beyond one year there was no difference between emergency and non-emergency patients in the proportion seen with a relevant symptom for either cancer.

Table 55 Percentage of patients with relevant consultation in the four years leading up to cancer diagnosis according to non-emergency and emergency routes

		Colon			Rectum		
		Elective	Emergency	p value	Elective	Emergency	p value
No Consultation within year		23.3	30.6	< 0.0005	21.8	38.8	< 0.0005
Consultation within 28days	0	71.3	52.1	< 0.0005	59.7	60.4	0.253
	1 - 2	27.6	42.0		39.6	35.1	
	3 - 4	1.1	5.4		0.7	4.2	
	5+	0.02	0.5		0.04	0.3	
Between 28 days to one year	0	38.1	56.7	< 0.0005	48.1	60.9	< 0.0005
	1 - 2	49.6	32.1		44.7	33.2	
	3 - 4	9.6	8.2		5.7	4.2	
	5+	2.7	3.0		1.5	1.6	
Between one to two years	0	81.5	81.1	0.256	85.3	83.1	0.237
	1 - 2	16.2	15.9		13.5	14.8	
	3 - 4	1.8	2.3		0.8	2.1	
	5+	0.5	0.7		0.4	0	
Between two to three years	0	84.8	85.5	0.843	88.7	87.6	0.275
	1 - 2	13.5	12.4		10.1	10.3	
	3 - 4	1.4	1.6		0.9	1.6	
	5+	0.3	0.5		0.2	0.5	
Between three to four years	0	87.6	87.9	0.929	89.5	86.3	0.079
	1 - 2	11.0	10.5		9.7	12.9	
	3 - 4	1.2	1.5		0.7	0.5	
	5+	0.2	0.5		0.1	0.3	

4.7.3.5 Symptoms before diagnosis

Tables 56 - 59

Rectal bleeding was more common in non-emergency patients compared with emergency patients for both cancer sites. Between 28 days to one year before cancer diagnosis, 12.4% of non-emergency colon and 21.8% of non-emergency rectal cancer patients presented with this symptom. This compares with 3.4% ($p < 0.0005$) and 6.1% ($p < 0.0005$) of emergency patients respectively. In the last 28 days before diagnosis, the percentage of rectal emergency patients with this symptom increased to 10.8%, but still remained lower than the 19% of elective patients. There was a drop in both colon emergency (6.3%) and non-emergency (1.9%) patients presenting with this symptom in the last 28 days before diagnosis ($p < 0.0005$).

There was no significant difference in the proportion of patients presenting with abdominal pain between emergency and elective patients, for both cancer sites, up to 28 days before diagnosis. However within 28 days, 28.6% of emergency colon patients presented with this symptom, as opposed to 7.3% elective. For rectal cancer, 9.8% of emergency patients presented with this symptom, as opposed to 2.4% of non emergency.

A change in bowel habit was more common for non-emergency patients for both the period between 28 days and year, and within 28 days, for both cancers. However the difference in the proportion of patients with this symptom was wider in the last 28 days between the two routes.

There was no significant difference in diarrhoea from 4 years up until the last 28 days. After this point there was a significantly higher proportion of those diagnosed as an emergency for both cancers (5.2% vs 2.8% for colon and 7.7% vs 5.1% for rectal).

There was no difference in the proportion of rectal cancer patients with constipation until the last 28 days before diagnosis, when 7.7% of emergency and 2.7% of elective patients complained of this symptom. For colon cancer, constipation was consistently a symptom seen higher in the emergency patients from two years before diagnosis.

Anaemia was noted consistently higher in the emergency group in all the time intervals before a rectal cancer diagnosis, but was significantly higher in the elective group for colon cancer only beyond a year leading up to diagnosis. Other bowel function only became significantly more common in emergency patients in the last 28 days for both cancers, as did weight loss or fatigue.

55.2% of non-emergency colon cancer patients compared with 30.7% of emergency patients presented with a red-flag symptom to their GP in the year before diagnosis ($p < 0.0005$). 21.3% of emergency colon cancer patients had previously presented to their GP with a red-flag symptom in between 28 days to a year before diagnosis. Non-emergency were more likely to present with a red-flag symptom in this period (41.3% $p < 0.0005$). In the 28 days before diagnosis, again the non-emergency patients were more likely to be seen with a red-flag symptom, but the difference was

less pronounced (19.9% vs 13.5% $p < 0.0005$). Of the emergency patients seen within 28 days to a year, 50.8% presented with only a non-specific symptom compared with 33.3% of non-emergency patients seen during this period ($p < 0.0005$).

For rectal cancer patients, 40.9% of emergency patients were previously seen by their GP with a red-flag symptom in between 28 days to a year before diagnosis, compared with 68.7% of non-emergency patients ($p < 0.0005$). This was also seen in the last 28 days (35.6% vs 21.9%) and between 28 days to one year (41.9% vs 25.9% $p < 0.0005$). Of all the emergency patients seen between 28 days to a year before diagnosis, 33.8% of emergency patients presented with only a non-specific symptom (33.8% vs 19.2% $p < 0.0005$).

Table 56 - Presence of symptoms within 28 days of cancer diagnosis

	Colon			Rectum		
	Elective	Emergency	p	Elective	Emergency	p
Abdominal pain	7.3	26.8	< 0.0005	2.4	9.8	< 0.0005
Anaemia	7.4	4.5	< 0.0005	1.9	3.7	0.028
Change in bowel habit	3.7	1.3	< 0.0005	10.5	1.8	< 0.0005
Constipation	1.9	9.5	< 0.0005	2.7	7.7	< 0.0005
Diarrhoea	2.8	5.2	< 0.0005	5.1	7.7	0.043
Fatigue	0.9	1.4	0.028	0.3	1.6	< 0.0005
Abdominal mass	0.8	1.6	0.002	0.7	0.5	0.688
Other bowel function	0.3	2.1	< 0.0005	0.4	1.6	0.003
Rectal Bleeding	6.3	1.9	< 0.0005	19	10.8	< 0.0005
Weight loss	1.0	1.9	< 0.005	1	2.1	0.043

Table 57 - Presence of symptoms between 28 days and one year before cancer diagnosis-

	Colon			Rectum		
	Elective	Emergency	p	Elective	Emergency	p
Abdominal pain	5.8	6.0	0.761	3.8	2.6	0.261
Anaemia	3.5	4.7	0.008	1.5	4.2	< 0.0005
Change in bowel habit	0.6	0.7	0.566	0.9	1.6	0.224
Constipation	2.6	4.1	< 0.0005	2.2	3.2	0.26
Diarrhoea	2.9	3.2	0.469	2.1	4	0.027
Fatigue	2.9	2.2	0.087	2.4	1.8	0.496
Abdominal mass	0.3	0.2	0.372	0.2	0	0.412
Other bowel function	0.7	0.7	0.900	0.7	0.1	0.518
Rectal Bleeding	2.2	1.3	0.014	2.5	1.8	0.427
Weight loss	0.8	0.9	0.588	0.5	0.8	0.458

Table 58 - Presence of symptoms between one and two years before cancer diagnosis

	Colon			Rectum		
	Elective	Emergency	p	Elective	Emergency	p
Abdominal pain	20.0	20.0	0.950	7.1	6.9	0.890
Anaemia	17.8	11.3	< 0.0005	4.5	5.8	0.244
Change in bowel habit	6.8	2.1	< 0.0005	9.6	4.7	0.002
Constipation	7.2	8.8	0.017	6.6	7.7	0.456
Diarrhoea	8.4	7.2	0.060	10.4	12.4	0.241
Fatigue	4.7	3.6	0.020	2.3	3.2	0.327
Abdominal mass	1.0	0.4	0.016	0.7	0.3	0.342
Other bowel function	2.3	1.7	0.091	2.4	1.8	0.496
Rectal Bleeding	12.4	3.4	< 0.0005	21.8	6.1	< 0.0005
Weight loss	2.8	2.4	0.295	1.7	2.6	0.179

Table 59 - Presence of symptoms between two and four years before cancer diagnosis

	Colon			Rectum		
	Elective	Emergency	p	Elective	Emergency	p
Abdominal pain	7.8	7.0	0.238	5.5	5	0.698
Anaemia	3.5	4.6	0.028	1.7	4.5	< 0.0005
Change in bowel habit	0.5	0.7	0.437	0.8	1.1	0.575
Constipation	4.5	5.3	0.102	3.0	4.5	0.126
Diarrhoea	4.5	4.6	0.891	4.0	3.2	0.446
Fatigue	4.3	3.8	0.244	2.9	5.5	0.006
Abdominal mass	0.4	0.2	0.423	0.5	0.3	0.534
Other bowel function	1.1	0.8	0.302	1.2	0.5	0.259
Rectal Bleeding	2.4	1.9	0.163	2.8	3.4	0.491
Weight loss	1.4	1.5	0.558	0.9	1.1	0.744

4.7.3.6 Logistic regression for an emergency diagnosis for all patients

In the binary logistic regression model (table 60), male patients were less likely to be diagnosed as an emergency for both colon (OR 0.85 95% CI 0.77 – 0.93) and rectal cancer (OR 0.76 95% CI 0.61 – 0.94 $p = 0.012$). Colon cancer patients under the age of 50 (OR 1.68 95% CI 1.46 – 1.94 $p < 0.0005$) and over the age of 80 (OR 1.68 95% CI 1.46 – 1.94 $p < 0.0005$) were more likely than those aged 60 – 69 years olds to be diagnosed as an emergency. Only patients aged over 80 with rectal cancer were more likely than the 60 – 69 year olds to be diagnosed as an emergency. Increasing deprivation also increased the likelihood of an emergency for both cancers (IMD 5 OR 1.56 95% CI 1.34 – 1.85 $p < 0.0005$) for colon, OR 1.97 95% CI 1.34 – 2.91 $p < 0.0005$ for rectal), as did increasing Charlson score.

Increasing number of consultations for any reason between 28 days to a year decrease the odds of an emergency diagnosis for both cancers. Colon cancer patients with anaemia (OR 0.50 95% CI 0.43 – 0.58 $p < 0.0005$), rectal bleeding (OR 0.26 95% CI 0.20 – 0.33 $p < 0.0005$), change in bowel habit (OR 0.32 95% CI 0.24 – 0.45 $p < 0.0005$) and diarrhoea (OR 0.82 95% CI 0.68 – 0.99 $p = 0.041$) between 28 days and a year before diagnosis were less likely to be diagnosed as an emergency. Presence of bleeding (OR 0.24 95% CI 0.16 – 0.37 $p < 0.0005$) and a change in bowel habit (OR 0.53 95% CI 0.33 – 0.89 $p = 0.009$) decreased the odds for a rectal cancer patient to be diagnosed as an emergency. No symptom individually was

associated with an increased odds of an emergency diagnosis. However patients who presented with only a non-specific symptom were significantly more likely to be diagnosed as an emergency for colon cancer (OR 1.15 95% CI 1.02 – 1.31 p = 0.026), but not rectal cancer.

Table 60 - The odds ratio for an emergency diagnosis according to patient demographics and presence or absence of symptoms between 28 days and a year before cancer diagnosis

		Colon				Rectum			
		OR	95% CI		p	OR	95% CI		p
Gender	Female	ref				ref			
	Male	0.85	0.77	0.93	0.001	0.76	0.61	0.94	0.012
Age	Under 50	1.68	1.46	1.94	< 0.0005	1.08	0.62	1.89	0.781
	50 - 59	1.04	0.86	1.14	0.675	0.74	0.46	1.19	0.213
	60 - 69	ref				ref			
	70 - 79	0.98	0.85	1.12	0.734	1.36	0.98	1.87	0.064
	80 and over	1.68	1.46	1.94	< 0.0005	2.22	1.63	3.09	< 0.0005
IMD	1	ref				ref			
	2	1.01	0.87	1.16	0.933	1.26	0.90	1.77	0.183
	3	1.20	1.03	1.39	0.02	1.44	1.00	2.07	0.048
	4	1.22	1.05	1.42	0.09	1.77	1.25	2.51	0.001
	5	1.56	1.34	1.85	< 0.0005	1.97	1.34	2.91	< 0.0005
Charlson	Zero	ref				ref			
	One	1.99	2.29	1.73	< 0.0005	3.52	2.69	4.64	< 0.0005
	Two or more	2.73	2.35	3.19	< 0.0005	4.21	3.02	5.89	< 0.0005
Consultations between 28 days and a year		0.99	0.98	0.99	< 0.0005	0.98	0.97	0.99	0.010
Symptoms									
Abdominal pain		1.11	0.98	1.26	0.111	1.00	0.62	1.61	0.997
Anaemia		0.50	0.43	0.58	<0.0005	0.80	0.51	1.28	0.353
Rectal Bleeding		0.26	0.20	0.33	< 0.0005	0.24	0.16	0.37	< 0.0005
Change in bowel habit		0.32	0.24	0.45	< 0.0005	0.53	0.33	0.85	0.009
Constipation		1.19	0.98	1.44	0.078	1.10	0.68	1.56	0.904
Diarrhoea		0.82	0.68	0.99	0.041	1.15	0.82	1.61	0.431
Mass		0.48	0.25	0.95	0.034	0.44	0.05	3.73	0.449
Other bowel function		0.83	0.59	1.18	0.298	0.86	0.38	1.92	0.704
Weight loss/fatigue		0.85	0.63	1.15	0.281	1.34	0.62	2.87	0.457
Only non-specific		1.15	1.02	1.31	0.026	1.38	0.98	1.96	0.067

4.7.3.7 Comparison of patients with missed opportunity for non-emergency diagnosis with those with an emergency

Between 2006 – 13, 4349 (56.7%) of colon and 1540 (51.0%) of rectal cancer patients had a recording of a relevant consultation between 28 days to one year (table 61). These were split according to emergency diagnosis as follows:

Table 61 – The number of colon and rectal cancer patients diagnosed as an emergency (study) and non emergency (control) with a symptom in between 28 days and year before cancer diagnosis

	Control		Study	
	n	%	n	%
Colon	3304	76.0	1045	24.0
Rectal	1396	90.6	144	9.4
Total	4700	79.8	1189	20.2

4.7.3.7.1 Patient demographics and tumour stage for colon cancer patients

Female patients were more likely to experience a missed opportunity for a non-emergency diagnosis in both univariate (25.4% vs 22.6%) and multivariate analysis (Male OR 0.85 95% CI 0.74 – 0.99 $p = 0.034$). Patients under the age of 50 (OR 2.16 95% CI 1.55 – 3.01 $p < 0.0005$) and those over the age of 80 (OR 1.62 95% CI 1.33 – 1.98 $p < 0.0005$) were at increased risk of a missed non-emergency diagnosis, as was increasing Charlson score, and a IMD of 5 (OR 1.46 95% CI 1.16 – 1.84 $p = 0.001$) (table 62). Patients who experienced a missed opportunity for a non-emergency diagnosis were more likely to be diagnosed at stage III (26.6% vs 26.4%) and IV (22.3% vs 15.3%), with higher missing stage (23.2% vs 14.9%) (table 64).

Table 62 - Univariate and multivariate analysis of demographics for colon cancer patients with a symptom recorded between 28 days to one year before cancer diagnosis

		Control		Study			OR	95% CI	p
		n	%	n	%				
Gender	Male	1671	77.4	489	22.6	0.033	0.85	0.74 - 0.99	0.034
	Female	1633	74.6	556	25.4		ref		
Age	Under 50	148	69.5	65	30.5	< 0.0005	2.16	1.55 - 3.01	< 0.0005
	50 - 59	334	79.3	87	20.7		1.25	0.91 - 1.71	0.174
	60 - 69	761	81.4	174	18.6		ref		
	70 - 79	1123	78.8	303	21.2		1.08	0.87 - 1.35	0.486
	80 and over	938	69.3	416	30.7		1.62	1.33 - 1.98	< 0.0005
Charlson score	Zero	2355	81.2	546	18.8	< 0.0005	ref		
	One	579	69.7	252	30.3		1.84	1.54 - 2.21	< 0.0005
	≥ 2	370	60	247	40		2.72	2.28 - 3.25	< 0.0005
IMD	1	822	78.2	229	21.8	0.001	ref		
	2	874	77.6	253	22.4		1.03	0.84 - 1.26	0.807
	3	659	76.5	202	23.5		1.07	0.86 - 1.34	0.55
	4	586	74.7	198	25.3		1.15	0.91 - 1.46	0.241
	5	363	69	163	31		1.46	1.16 - 1.84	0.001

For rectal cancer (table 63), age 80 and over was associated with a significantly more likely to experience a missed opportunity in primary care before an emergency in both univariate (16.9%) and multivariate analysis (OR 2.93 95% CI 1.72 – 4.99 $p < 0.0005$). Again increasing Charlson score was associated with a missed opportunity for a non-emergency diagnosis. Gender, IMD or young age made no significant difference to the likelihood of an emergency diagnosis. A higher proportion of patients from the study cohort were diagnosed with stage III (22.2%), stage IV (23.6%) and missing stage cancer (38.2%) than either stage I (6.9%) or stage II (9%).

Table 63 – Univariate and multivariate analysis of demographics for rectal cancer patients with a symptom recorded between 28 days to one year before diagnosis

		Non-emergency		Emergency		p	OR	95% CI	p
		n	%	n	%				
Gender	Male	856	91.2	83	8.8	0.389	0.84	0.60 - 1.18	0.314
	Female	540	89.9	61	10.1		ref		
Age	Under 50	99	93.4	7	6.6	< 0.0005	1.47	0.62 - 3.51	0.385
	50 - 59	221	95.3	11	4.7		1	0.45 - 2.21	0.999
	60 - 69	326	94.5	19	5.5		ref		
	70 - 79	430	91.1	42	8.9		1.54	0.85 - 2.79	0.157
	80 and over	320	83.1	65	16.9		2.93	1.72 - 4.99	< 0.0005
Charlson score	Zero	1089	93.8	72	6.2	< 0.0005	ref		
	One	195	84.1	37	15.9		2.51	1.67 - 3.78	< 0.0005
	Two or more	112	76.2	35	23.8		3.93	2.49 - 6.22	< 0.0005
IMD	1	320	92.5	26	7.5	0.436	ref		
	2	342	91.2	33	8.8		1.14	0.69 - 1.89	0.611
	3	294	91	29	9		1.09	0.62 - 1.90	0.769
	4	255	88.9	32	11.1		1.37	0.80 - 2.36	0.252
	5	185	88.5	24	11.5		1.28	0.71 - 2.32	0.416

Table 64 - Stage at diagnosis for colon and rectal cancer patients for those who had a relevant symptom recorded between 28 days to a year before diagnosis and were diagnosed as a non-emergency (control) and those that had an emergency diagnosis (study)

		Colon				Rectum			
		Control		Study		Control		Study	
		n	%	n	%	n	%	n	%
Stage	I	375	11.3	33	3.2	236	16.9	10	6.9
	II	1057	32	259	24.8	251	18	13	9
	III	873	26.4	278	26.6	398	28.5	32	22.2
	IV	506	15.3	233	22.3	178	12.8	34	23.6
	Missing	493	14.9	242	23.2	333	23.9	55	38.2

Chi square analysis demonstrated a significant difference of < 0.0005 for both cancers

4.7.3.7.2 Symptoms associated with a missed diagnosis

Colon cancer patients presenting with rectal bleeding (OR 0.36 95% CI 0.28 – 0.46 $p < 0.0005$), anaemia (OR 0.68 95% CI 0.58 – 0.80 $p < 0.0005$) or a change in bowel habit (OR 0.46 95% CI 0.33 – 0.64 $p < 0.0005$) were more likely to be identified by the GP, leading to a non-emergency diagnosis in both univariate and multivariate analysis (table 65). Presenting with abdominal pain (OR 2.10 95% CI 1.80 – 2.44 95% CI 1.80 – 2.44 $p < 0.0005$) or constipation (OR 1.78 95% CI 1.45 – 2.18 $p < 0.0005$) between 28 days and to a year before diagnosis increased the likelihood of a missed opportunity of a non-emergency diagnosis. No other symptom reached significance on multivariate analysis although the presence of diarrhoea was significantly associated with a missed diagnosis on univariate analysis. The presence of only non-specific symptoms increased the odds ratio of a missed diagnosis (OR 2.34 95% CI 2.00 – 2.74 $p < 0.0005$).

Table 65 - Symptom before diagnosis for colon cancer patients and OR of the presence of that symptom leading to an emergency diagnosis

		Control		Study		p	OR	95% CI	p
		n	%	n	%				
Red-flag	Yes	2229	81.1	518	18.9	< 0.0005		ref	
	No	1075	67.1	527	32.9		2.34	2.00 - 2.74	< 0.0005
Rectal bleeding	No	2640	73.3	962	26.7	< 0.0005		ref	
	Yes	664	88.9	83	11.1		0.36	0.28 - 0.46	< 0.0005
Anaemia	No	2325	75.1	771	24.9	0.034		ref	
	Yes	979	78.1	274	21.9		0.68	0.58 - 0.80	< 0.0005
Change in bowel habit	No	2938	74.7	994	25.3	< 0.0005		ref	
	Yes	366	87.8	51	12.2		0.46	0.33 - 0.64	< 0.0005
Diarrhoea	No	2858	76.7	870	23.3	0.009		ref	
	Yes	446	71.8	175	28.2		1.21	0.98 - 1.49	0.07
Mass	No	3253	75.9	1035	24.1	0.16		ref	
	Yes	51	83.6	10	16.4		0.65	0.32 - 1.32	0.237
Abdominal pain	No	2246	80	563	20	< 0.0005		ref	
	Yes	1058	68.7	482	31.3		2.1	1.80 - 2.44	< 0.0005
Constipation	No	2928	77.8	836	22.2	< 0.0005		ref	
	Yes	376	64.3	209	35.7		1.78	1.45 - 2.18	< 0.0005
Other bowel function	No	3179	76	1003	24	0.729		ref	
	Yes	125	74.9	42	25.1		1.17	0.81 - 1.68	0.401
Weight loss or fatigue	No	2910	76.3	904	23.7	0.179		ref	
	Yes	394	73.6	141	26.4		1.06	0.87 - 1.29	0.574

Bleeding (OR 0.27 95% CI 0.17 – 0.42 $p < 0.0005$) was the only symptom associated with a non-emergency diagnosis for rectal cancer patients, while diarrhoea was associated with a missed diagnosis (OR 1.77 95% CI 1.22 – 2.59 $p = 0.003$), as was the presence of only non-specific symptoms (OR 2.29 95% CI 1.56 – 3.38 $p < 0.0005$) (table 66). Although anaemia, constipation and weight loss or fatigue were more common in the missed diagnosis group on univariate analysis, the odds of a missed diagnosis did not reach significance on multivariate analysis.

Table 66 - Symptom before diagnosis for rectal cancer patients and OR of the presence of that symptom leading to an emergency diagnosis

		Non-emergency		Emergency					
		n	%	n	%	p	OR	95% CI	p
Red-flag	Yes	1129	92.3	94	7.7	< 0.0005	Ref		
	No	267	84.2	50	15.8		2.29	1.56 - 3.38	< 0.0005
Rectal bleeding	No	804	86.8	122	13.2	< 0.0005	Ref		
	Yes	592	96.4	22	3.6		0.27	0.17 - 0.42	< 0.0005
Anaemia	No	1276	91.1	124	8.9	0.035	Ref		
	Yes	120	85.7	20	14.3		1.02	0.60 - 1.73	0.947
Change in bowel habit	No	1140	90	127	10	0.51	Ref		
	Yes	256	93.8	17	6.2		0.73	0.44 - 1.22	0.224
Diarrhoea	No	1116	91.9	98	8.1	0.001	Ref		
	Yes	280	85.9	46	14.1		1.77	1.22 - 2.59	0.003
Mass	No	1377	90.6	143	9.4	0.501	Ref		
	Yes	19	95	1	5		0.55	0.07 - 4.45	0.572
Abdominal pain	No	1208	91.1	118	8.9	0.13	Ref		
	Yes	188	87.9	26	12.1		1.56	0.94 - 2.56	0.083
Constipation	No	1222	91.4	115	8.6	0.01	Ref		
	Yes	174	85.7	29	14.3		1.56	0.96 – 2.52	0.054
Other bowel function	No	1333	90.7	137	9.3	0.849	Ref		
	Yes	63	90	7	10		1.28	0.57 - 2.90	0.552
Weight loss or fatigue	No	1292	91.2	125	8.8	0.015	ref		
	Yes	104	84.6	19	15.4		1.61	0.93 - 2.79	0.09

Patients with a missed diagnosis for colon cancer demonstrated a worse stage at diagnosis compared to non-emergency colon cancer patients for all symptoms, except for rectal bleeding. For example the presence of abdominal pain was associated with an OR of 1.50 ($p = 0.001$) of later stage of diagnosis for patients who experienced a missed diagnosis compared with the same symptom for those with a non-emergency diagnosis (table 67). A similar finding was noted if all red-flag symptoms or only non-specific symptoms were analysed together. A worse stage at diagnosis was noted for rectal bleeding for rectal cancer patients, or for red-flag and non-specific as a group.

Table 67 - Odds ratio of a late stage (III, IV or missing) for patients with a missed diagnosis according to symptom compared with those diagnosed as a non-emergency

	Colon		Rectum	
	OR (95% CI)	p	OR (95% CI)	p
Red-flag	2.08 (1.68 – 2.57)	< 0.0005	2.61 (1.46 – 4.65)	0.001
Non-specific only	1.59 (1.23 – 2.04)	< 0.0005	2.45 (1.01 – 5.91)	< 0.0005
Rectal bleeding	1.50 (0.91 – 2.47)	0.12	3.7 (1.07 – 12.8)	0.04
Anaemia	2.00 (1.49 – 2.68)	< 0.0005	2.18 (0.59 – 8.08)	0.25
Abdominal pain	1.50 (1.18 – 1.92)	0.001	1.38 (0.46 – 4.17)	0.567
Constipation	1.91 (1.29 – 2.85)	0.001	2.70 (0.86 – 8.43)	0.09

4.7.4 Discussion

4.7.4.1 Summary

The aim of this study was to determine whether there were differences in patient demographics, GP attendance and symptoms between non-emergency and emergency colorectal cancer patients. The second aim was to determine the demographics and symptoms of emergency patients who had previously consulted their GP with cancer related symptoms leading up to their diagnosis. By using a large database, representative of the English cancer population, this study demonstrates patients those with non-specific symptoms, females, lower deprivation and the extreme ends of age, are more likely to be diagnosed as an emergency, while rectal bleeding and other 'red-flag' symptoms seem to be protective. Colon cancer was associated with a higher emergency diagnosis rate than rectal cancer.

As discussed in the introduction, patients who present as an emergency do so for many reasons. These can be broadly differentiated as modifiable versus non-modifiable⁹³. An example of the latter are tumours which first present themselves with severe symptoms, such as perforation, thereby requiring the patient to bypass the GP and seek immediate medical attention in the emergency department. On the contrary, GP recognition of patients with underlying cancer dependent on symptoms is potentially modifiable through national policy. This study provides a target for such improvement, in particular the 21.3% of emergency colon cancer and 25.9% of rectal cancer patients who had previously presented to primary care in the 28 days to a year before diagnosis with a red-flag symptom, but yet still were not identified as a possible cancer patient.

Patients diagnosed through screening, DCO or the 'unknown route' were excluded, so the remaining non-emergency patients, of which TWW (34.1% of all cancers) were the most common for both cancers, would have presented to their GP at least once in the year before the cancer diagnosis. Despite this, in the first part of the study 23.3% of colon non-emergency and 21.8% of rectal non-emergency patients did not have a relevant symptom recorded in CPRD in the year before diagnosis. Therefore the 30.6% of the colon and 38.8% of rectal emergency patients who did not have a symptom recorded do not simply represent

the population who bypass their GP and present directly to Accident and Emergency. Both emergency and non-emergency patients had a similar proportion of patients with at least one recording of a consultation for any reason in the year leading up to diagnosis, indicating there is not an inherent bias against emergency patients for missing data. Instead the missing data is more likely due to the type of symptoms. In a study by Price et al.¹²⁴ the free text of consultations was compared to the medcodes in CPRD. The authors found symptoms considered to be non-specific were less likely to be found within the medcodes compared to alarm symptoms. Although the free text is no longer available to researchers using CPRD, it can be assumed missing consultations are more likely to be due to non-specific symptoms. Therefore inclusion of the missing symptoms is likely to add weight to our findings. Other reasons for missing data are discussed in more detail later.

An important finding of the first part of the study is that there is no difference in the number or rate of consultations for any reason, or symptom related consultations, between emergency and non-emergency patients beyond one year before diagnosis. If the rate of emergency diagnoses is to be reduced, the opportunity to do so will be found within a year before diagnosis.

Of the rectal cancers, 12.3% were diagnosed as an emergency, as opposed to 31.5% of colon cancers, while 45.1% and 29.7% of each cancer was diagnosed via the TWW route. This corresponds to the higher proportion of rectal cancers who had rectal bleeding (21.8%) compared with colon cancer (12.4%) in the period between 28 days and a year before a non-emergency diagnosis. Rectal cancers were also less likely to present with constipation, and more likely to present with a change in bowel habit during this period. These findings suggest symptoms from rectal cancers are more obvious clinically due to the proximity to the anus, and therefore are less likely to be diagnosed as an emergency. This corresponds to previous findings that a higher proportion of “hidden” cancers, such as brain and leukaemias⁵⁵, are diagnosed as an emergency or later stage⁸⁸.

The second part of this study is the first to define the group of patients who experienced an emergency diagnosis despite presenting to the GP. In further analysis of the Route of diagnosis algorithm by Elliss-Brookes et al.⁹⁸ it has been suggested 56% of colorectal cancer patients who are diagnosed as an emergency attend Accident and Emergency directly without a referral from their GP. However this result does not define the cohort who saw their GP before attending the emergency department, and so experienced a missed opportunity for a non-emergency diagnosis. The findings in second part of the study are similar to the first in terms of the demographics associated with an emergency diagnosis, such as the extreme ends of age and a higher Charlson score. However the second part demonstrated the non-specific symptoms that were associated with a missed opportunity for a non-emergency, such as abdominal pain and constipation.

As expected emergency patients, whether colon or rectum, were diagnosed at a later stage. Rectal cancers also had the higher stage I cancers compared to colon cancers, for both non-emergency (17.3% vs 10.8%) and emergency (8.2% vs 2.6%) patients. This again may be because rectal cancer symptoms, such as rectal bleeding, are more apparent and are therefore detected sooner by the GP. A novel finding of this study is that emergency patients were diagnosed at a later stage compared with non-emergency patients despite presenting with the same symptoms. One explanation is that emergency patients have particularly aggressive tumours that are more likely to be diagnosed at a later stage. Another explanation is the duration of symptoms before presentation. As mentioned in the introduction, certain patients may wait before seeking medical attention due to the fear of diagnosis. During this time the stage will progress, and present with symptoms such as obstruction, requiring the patient to bypass the GP. Unfortunately the medcodes in CPRD only describe the presence of the symptom rather than its severity or duration. Within the group of patients with a missed diagnosis, there will be a cohort who presented to accident and emergency while they were waiting for a hospital appointment or results. It is difficult to determine in CPRD the proportion of patients who are diagnosed through this route.

4.7.4.2 Comparison to previous studies

This is the first study to define a cohort of emergency patients for whom there was an opportunity for a non-emergency diagnosis. This is the largest such investigation, from a cohort that is representative of the general cancer population. There is a wide variation in the definition of emergency patients in published studies, which makes comparisons difficult. Some studies have defined emergency patients as those who have had an emergency operation, rather than presenting to an emergency department. Such a definition would exclude the elderly patients, with disseminated cancer, for whom an operation is not an option, but who nevertheless could have experienced missed opportunities for an earlier diagnosis. The reason for this definition lies in the difficulty in obtaining both inpatient and primary care notes for patients on a large scale, a combination which is available in CPRD.

Two previous studies have attempted to define the GP usage before an emergency diagnosis by using a combination of primary and secondary care records. Renzi et al.¹⁰⁵ used CPRD to identify 1606 colorectal cancer patients between 2005 – 06. Symptom medcodes were similar to my study. They also found there was no difference in consultation numbers beyond a year before diagnosis. However in my study there were significant differences in the proportion of patients with rectal bleeding, constipation and a change in bowel habit within the last 28 days and up to a year, and in the odds ratio of an emergency for a change in bowel habit for rectal cancers and diarrhoea for colon cancers which were not identified in their study. The discrepancies could be explained by the six fold increase in patient numbers included in my study, and that they defined rectal cancer as ICD C19 – C20 as opposed to my definition (and that of the national cancer registry definition) of C20. Sheringham et al.⁹⁷ obtained records from a region within England for 943 patients between 2009 - 11, but analysed all consultations up to diagnosis date, rather than excluding the last 28 days. Without the latter it is difficult to define patients who had experienced a missed opportunity for an emergency diagnosis. Due to the lower numbers they did not carry out multivariate analysis for all symptoms, instead concentrating on anaemia, bleeding, constipation and pain. Also colon and rectal cancers were analysed together. They

concluded anaemia bleeding was associated with a non-emergency diagnosis, while constipation and pain was associated with an emergency diagnosis.

4.7.4.3 Limitations

It is not possible to be certain that the symptom experienced leading up to cancer diagnosis was due to the underlying cancer, such as rectal bleeding could be from hemorrhoids rather than cancer. This is a limitation with any retrospective study that investigates symptomology. Anaemia, which is more common in elderly patients with multiple benign co-morbidities, was associated with a reduced likelihood of an emergency presentation, suggesting GPs are not assuming the anaemia is due to an underlying benign condition.

When a GP makes an emergency referral to secondary care for the symptoms described here, the vast majority of patients would have seen medical practitioners in secondary care on the same day. It is likely the diagnosis would have been made during that visit, or shortly after in the community. The 28 day cut-off used in this study covers this pathway, and so assumes consultations during this period have directly led to the cancer diagnosis. These are not missed opportunities for diagnosis. Although this is a definition which has previously been used, it is difficult to confirm this reflects the pathway for all patients. Of note there was a significant difference in the proportion of patients presenting to their GP before and after the 28 day limit with a particular symptom, indicating this threshold is able to identify differences in GP usage along the diagnostic pathway.

It is possible the emergency patients who had presented to their GP, had not experienced the symptoms for a sufficient time to fulfil the referral criteria. The duration of symptoms is not available in CPRD so it is not possible to determine the percentage of patients this applies to. Two systematic reviews identified the type of symptom and co-morbidities had the greatest influence on the length of time a patient would wait before presenting to primary care. Both of these have been included within the regression models.

Limitations regarding missing clinical and stage data and the route of diagnosis are applicable to all studies in this thesis, and so is discussed in more detail later in the main discussion. Briefly in reference to this chapter there is an assumption that all non-emergency routes would have had GP involvement, and so have been amalgamated in previous studies. However it is possible 'inpatient elective' or 'other outpatients' patients were diagnosed without GP involvement, although the proportion is unknown.

4.7.4.4 Implications for clinical practice

This study suggests patients falling outside the NICE referral criteria, due to a combination of age and symptom, are more likely to be diagnosed as an emergency. These referral criteria were implemented initially in 2000, and then updated in 2005 and again in 2015, with the intention to improve the identification of cancer patients in primary care, and by proxy, reduce emergency diagnoses. However despite the introduction, the proportion of emergency diagnoses has not dropped massively. The 2005 iteration was based on studies but did not state a positive predictive value threshold for the referral criteria. The latest guidelines have been based on a minimum PPV of 3%. As such, these guidelines have removed the requirement for the duration of symptoms and have introduced referral criteria for those under the age of 50, albeit a combination of symptoms rather than a single symptom as in older patients. It remains to be seen whether this will reduce the emergency diagnosis rate, in particular for the fifth of colon and quarter of rectal cancer emergency patients who presented with a red-flag symptom but yet still were diagnosed as an emergency. Reducing the threshold for referral may reduce the number of patients diagnosed as an emergency, but at the cost of increasing the use of resources in secondary care. Symptoms falling outside the referral criteria associated with a worse outcome is a recurrent theme, and will be discussed in more detail in the main conclusion chapter.

4.7.4.5 Alternative method of study

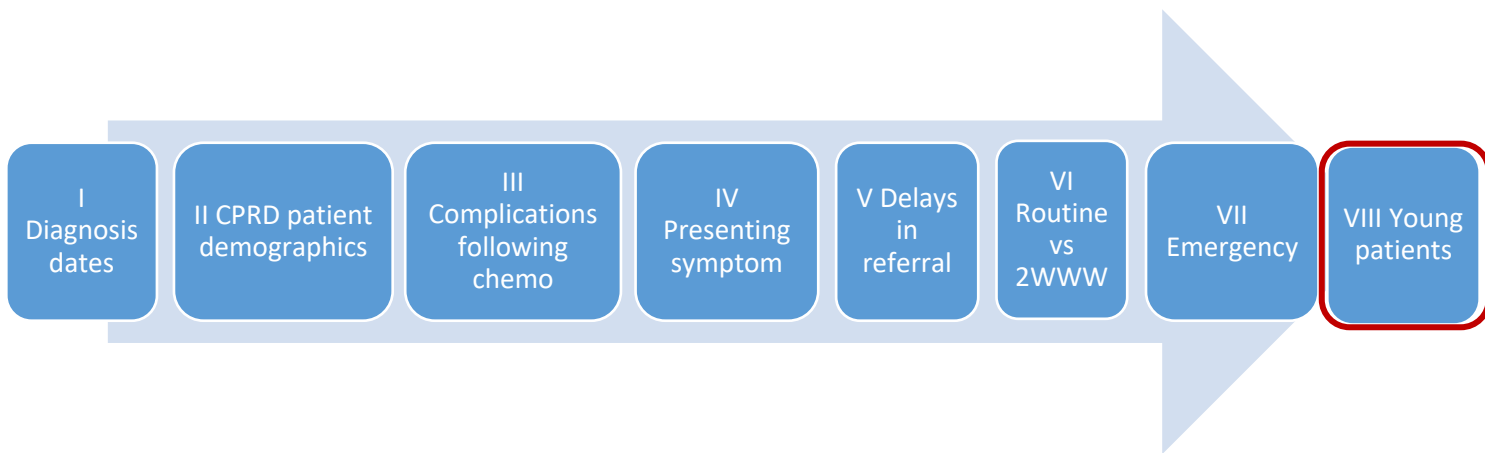
CPRD has provided a unique opportunity to explore the differences between non-emergency and emergency patients, on a scale that is not available in other datasets, and even less possible with retrospective notes analysis. QResearch is an alternative dataset but

does not provide the coverage or the linkage to the route of diagnosis provided by CPRD. Since I acquired the dataset used in this thesis, the route of diagnosis algorithm has expanded further to include differentiation of emergency patients into those that present via the GP or directly to the emergency department. In addition, the completeness of CPRD has improved over the years. This detail, together with less missing data, it may be possible to further clarify whether patients saw their GP before an emergency diagnosis or bypassed primary care completely.

4.7.4.6 Relevance to thesis

This chapter describes a potential area of change that could lead to an improvement in cancer care within England, and the UK, by reducing the proportion of cancers that are diagnosed as an emergency. This continues the theme from the previous chapter that non-specific symptoms are associated with a worse prognosis, while red-flag symptoms provide an opportunity for a prompt diagnosis. In the next chapter I explore the possibility of improving outcomes for younger patients.

End of Study



4.8 Study VIII – Do young patients experience delays in primary care?

In the previous chapter I identified patient symptoms and demographics associated with missed opportunities in primary care for non-emergency cancer diagnoses. In this study I concentrate on young patients, and the missed opportunity for an earlier diagnoses in this age group.

The hypothesis states:

- **Young patients experienced missed opportunities for a prompt referral even if they present with similar symptom to older patients**

The research questions for this study are:

- **Are there any differences in the presenting symptom between young patients, defined as under 50, and the older age groups?**
- **What is the difference in the primary, hospital and diagnostic interval between the age groups?**
- **Are young patients more likely to be diagnosed as an emergency even if they present with the same symptoms as older patients?**
- **Is there a difference in stage at diagnosis between young and older patients despite presentation with similar symptoms?**

4.8.1 Introduction

The incidence of colorectal cancer in young patients is increasing¹⁷⁹, with a suggestion that they are more likely to present late and/or as an emergency¹⁸⁰. However the national referral guidelines are primarily designed to identify patients over the age of 50. By better understanding the difference in presenting symptoms between young and older patients, there is the potential to improve their prognosis.

4.8.2 Methods

4.8.2.1 Diagnostic route

As explained earlier, the route of diagnosis was defined according to the protocol by Elliss-Brookes et al.²⁹ for all diagnoses after 2006 inclusive, therefore all earlier cases were excluded. 'Two-week wait', 'urgent', 'in-patient elective' and 'other outpatient' cases were considered non-emergency, while DCO, screening and 'unknown' were excluded as it is unclear whether the patient would have previously presented to their GP.

4.8.2.2 Presenting symptom

The first recording of a relevant symptom in the year leading up to the cancer registry diagnosis date, excluding the final 28 days, was considered to be the first CRC-associated presentation to the GP. As defined by Renzi et al.¹⁰⁵, consultations in the final 28 could have led directly to the diagnosis, and so do not represent missed opportunities to identify a patient with cancer. The symptom recorded at the first consultation was compared with age in univariate analysis individually and then as red-flag (bleeding, anaemia, change in bowel habit, diarrhoea, abdominal mass) or only non-specific (constipation, abdominal pain, weight loss, fatigue). Logistic regression modelling was used to determine the odds ratio of each age group to present with a particular symptom, with reference to the young group. This model was adjusted a priori for patient demographics, site of tumour and emergency diagnosis.

4.8.2.3 Consultations leading up to diagnosis

The total number of consultations related to associated symptoms were compared between the age groups. This was repeated after separating consultations into red-flag and non-specific.

The average consultation rate per patient per month for associated symptoms in the twelve months leading to diagnosis was used to determine whether there was a difference in health seeking behaviour between the age groups. This was further stratified according to presentation (emergency vs non-emergency).

4.8.2.4 Symptom leading to emergency diagnosis

The OR of an emergency diagnosis for each age group compared with the young patients for a particular symptom was calculated in a logistic model, adjusted a priori for patient demographics. For these models, only patients with the symptom at presentation of interest were included.

4.8.2.5 Interval to referral and the number of consultations

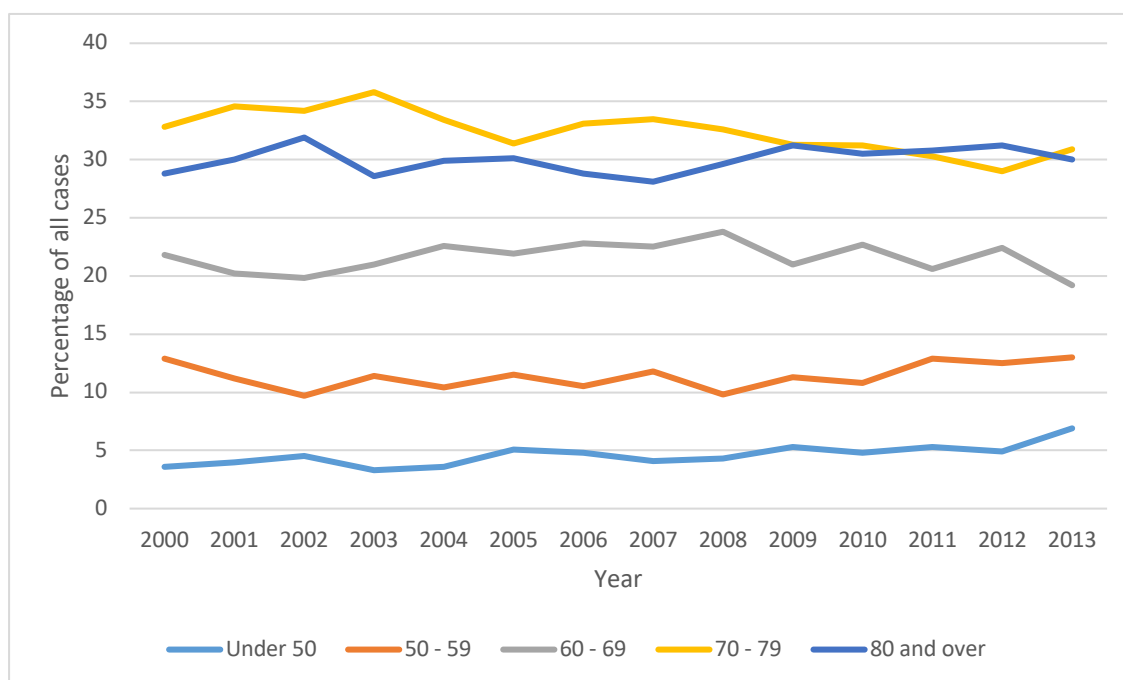
From the initial study cohort, patients with a recording of a relevant symptom in the year leading up to diagnosis, except for the last 28 days, and a referral to secondary care for cancer, were identified to compare the referral interval (presentation to referral), hospital interval (from referral to diagnosis) and diagnostic interval (from presentation to diagnosis).

4.8.3 Results

4.8.3.1 Change in diagnosis rates between 2000 – 2013

Between 2000 and 2013, 22,019 patients were identified as having a diagnosis of CRC recorded in the cancer registry. The largest rise in CRC cases was seen in the youngest age group, from 3.6% in 2000 to 6.9% in 2013, while a smaller rise was noted in those aged 80 and over (28.8% to 30%). Over the same period, the percentage of cases fell in those aged 60 – 69 (21.8% vs 19.2%) and 70 – 79 (32.8% vs 30.9%) (figure 30).

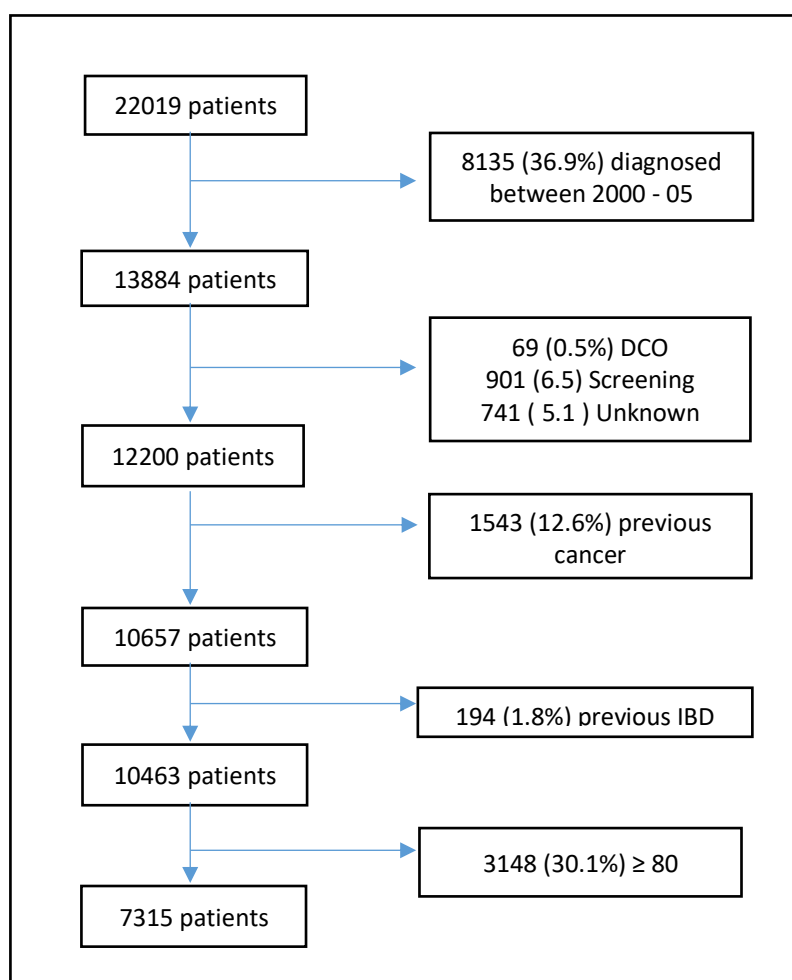
Figure 30 - Proportion of colorectal cancer diagnoses by age per year from 2000-2013 inclusive



4.8.3.2 Study cohort inclusion

Between 2000 and 2005 8135 (36.9%) patients were diagnosed with CRC. A further 3227 patients were excluded due to screening, DCO, unknown diagnostic route and a previous diagnosis of cancer or underlying IBD (figure 31). Of the remaining 10463 patients, 561 (5.0%) were under the age of 50, 1287 (11.5%) between 50 – 59, 2458 (21.9%) between 60 – 69, 3534 (31.5%) between 70 – 79 and 3478 (30.9%) 80 years and over. After excluding the oldest group, 7315 patients formed the study cohort.

Figure 31 - Study patient selection. DCO: Death certificate only IBD: Inflammatory bowel disease



4.8.3.3 Patient demographics and stage of tumour

A significantly higher proportion of patients under the age of 50 were female (51.4%) compared to all other age groups (table 68). The Charlson score increased with age, with 53.5% of 70 -79 year olds scored zero compared to 89.6% of those under the age of 50 ($p < 0.0005$). There was no significant difference in IMD ($p = 0.552$). The youngest age group had the highest proportion of patients with an emergency diagnosis (29.1%), and the lowest through a two week wait referral (22.6%) ($p < 0.0005$). The most common route for this group was the non-urgent 'GP referral' (33.3%). Patients aged 60 – 69 had the highest proportion of two week wait referrals (38.5%). Early stage cancer was significantly less common in patients under the age of 50, compared with all other age groups (31.8% vs 36.1%, 35.1% and 39.3% respectively), while missing stage was highest in this group, (20.5% vs 13.0%, 16.0% and 18.8%). The youngest age group had the highest proportion of late stage disease (68.5% vs 63.0%, 64.4%, 60.4% $p < 0.0005$) after missing was included as late stage.

Table 68 - Patients demographics, stage and site of colorectal cancer

		Under 50		50 - 59		60 - 69		70 - 79		All		p value
		n	%	n	%	n	%	n	%	n	%	
Gender	Male	247 _a	48.6	666 _b	57.0	1401	61.1	1918 _b	57.3	4232	57.9	p < 0.0005
	Female	261 _a	51.4	502 _b	43.0	893	38.9	1427 _b	42.7	3083	42.1	
Charlson	0	455 _a	89.6	932 _b	79.8	1566 _c	68.3	1790 _d	53.5	4743	64.8	p < 0.0005
	1	45 _a	8.9	166 _b	14.2	449 _c	19.6	770 _d	23.0	1430	19.5	
	2+	8 _a	1.6	70 _b	6.0	279 _c	12.2	785 _d	23.5	1142	15.6	
IMD	1	116	22.8	282	24.1	500	21.8	778	23.3	1676	22.9	p = 0.552
	2	128	25.2	285	24.4	565	24.6	867	25.3	1825	24.9	
	3	92	18.1	242	20.7	483	21.1	673	20.1	1490	20.4	
	4	96	18.9	194	16.6	422	18.4	626	18.7	1338	18.3	
	5	76	15.0	165	14.1	324	14.1	421	12.6	986	13.5	
Presentation	TWW	115 _a	22.6	435 _b	27.2	883 _b	38.5	1230 _b	36.8	2663	36.3	p < 0.0005
	GP referral	169 _a	33.3	331 _b	28.3	608 _b	26.5	895 _b	26.8	2003	27.4	
	Other outpatient	38 _a	7.5	103 _a	8.8	179 _a	7.8	290 _a	8.7	610	8.3	
	Inpatient elective	38	7.5	67 _{a,b}	5.7	146 _{a,b}	6.4	172 _b	5.1	423	5.8	
	Emergency	148 _a	29.1	232 _b	19.9	478 _{b,c}	20.8	758 _c	22.7	1616	22.1	
	Non-emergency	360 _a	70.9	936 _b	80.1	1816 _{b,c}	79.2	2587 _c	77.3	5699	77.9	
Stage	Early	160 _a	31.5	432 _{b,c}	37.0	817 _{a,c}	35.6	1323 _b	39.6	2732	37.3	p < 0.0005
	Late	244 _a	48.0	584 _a	50.0	1111 _a	48.4	1393 _b	41.6	3332	45.6	
	Missing	104 _a	20.5	152 _b	13.0	366 _c	16.0	629 _a	18.8	1251	17.1	
Site	Colon	343 _{a,b}	67.5	739 _b	63.3	1601 _a	69.8	2416 _c	72.2	5099	69.7	p < 0.0005
	Rectum	165 _{a,b}	32.5	429 _b	36.7	693 _a	30.2	929 _c	27.8	2216	30.3	

Similar letter subscripts indicate no significance difference between groups, with significance set at < 0.05

4.8.3.4 Number of consultations before diagnosis

A similar proportion of young patients were seen at least once due to an associated symptom in the year leading up to cancer diagnosis compared with all age groups except the 50 – 59 year olds (table 69). 3.3% of young patients were seen by their GP five or more times before diagnosis, which was

higher than all other ages. The young group had the highest proportion of patients seen at least once with a non-specific symptom (48.4% vs 27.4%, 27.4%, 28.2% respectively), although the proportion of young patients seen at least once with a red-flag symptom was similar to all other age groups. The proportion of patients seen three or more times did not differ with age after separating into red-flag or non-specific.

Table 69 - Number of consultations in the year leading up to the cancer diagnosis excluding the last 28 days

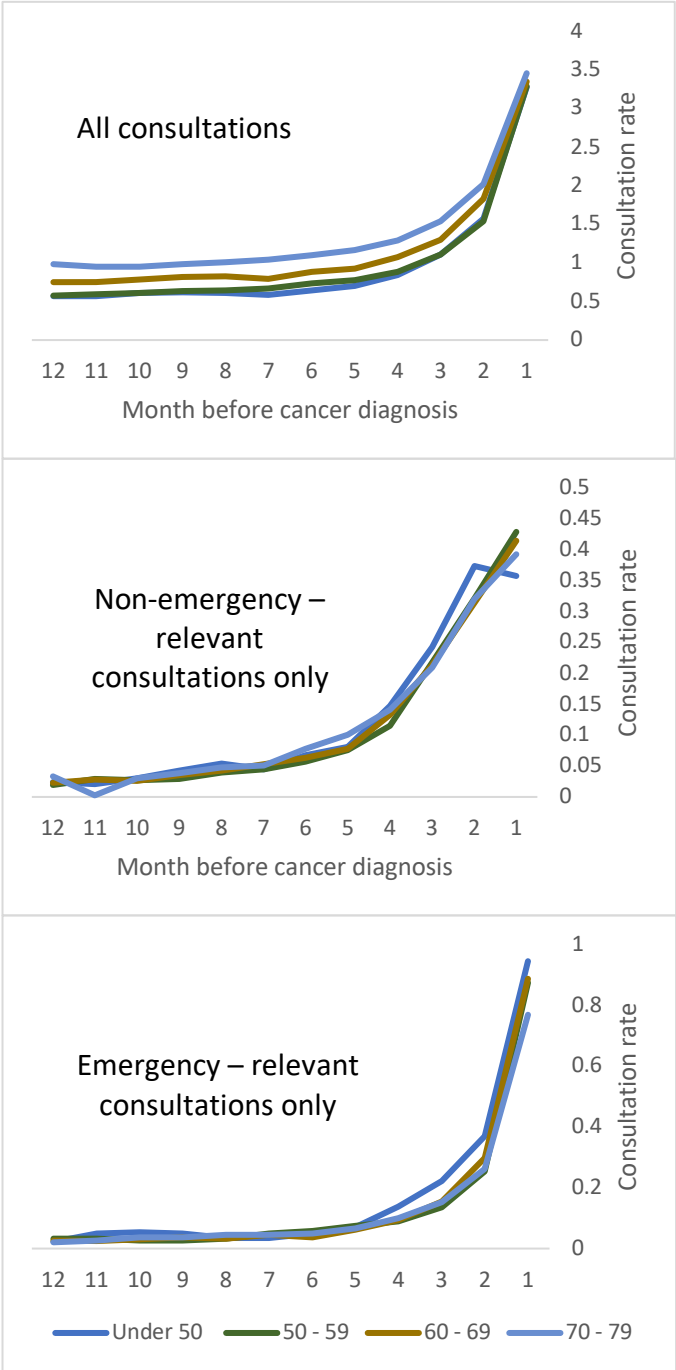
	Frequency	Under 50		50 - 59		60 - 69		70 - 79		All		
		n	%	n	%	n	%	n	%	n	%	
Relevant consultations	0	203 _a	40.0	530 _b	45.4	1022 _{a,b}	44.6	1448 _{a,b}	43.3	3203	43.8	p = 0.029
	1 – 2	243 _a	47.8	521 _a	44.6	1050 _a	45.8	1513 _a	45.2	3327	45.5	
	3 – 4	45 _a	8.9	99 _a	8.5	180 _a	7.8	300 _a	9.0	624	8.5	
	5+	17 _a	3.3	18 _b	1.5	42 _b	1.8	84 _b	2.5	161	2.2	
Consultations with red-flag symptoms	0	333 _{a,b}	65.6	754 _b	64.6	1464 _{a,b}	63.8	2051 _a	61.3	4928	62.9	p = 0.036
	1 – 2	156 _a	30.7	383 _a	32.8	766 _a	33.4	1157 _a	33.7	2462	33.7	
	3 – 4	17 _{a,b}	3.3	26 _b	2.2	61 _{a,b}	2.7	119 _a	3.0	223	3.0	
	5+	2 _{a,b}	0.4	5 _{a,b}	0.4	3 _b	0.1	18 _a	0.4	28	0.4	
Consultations with non-specific symptoms	0	318 _a	62.6	848 _b	72.6	1666 _b	72.6	2403 _b	71.8	5235	71.6	p = 0.003
	1 - 2	160 _a	31.5	262 _b	22.4	520 _b	22.7	788 _b	23.6	1730	23.7	
	3 - 4	22 _a	4.3	49 _a	4.2	85 _a	3.7	117 _a	3.5	273	3.7	
	5+	8 _a	1.6	9 _a	0.8	23 _a	1.0	37 _a	1.1	77	1.1	

Similar letter subscripts indicate no significance difference between groups, with significance set at < 0.05

4.8.3.5 Consultation rate in the year leading up to diagnosis

The consultation rate for all type consultations was highest for the oldest age group, and decreased with each age group (figure 32). For non-emergency patients (figure 32), there was no difference in the relevant consultation rate between the age groups. However, the rate increased for young patients earlier (at five months) compared with the other age groups before an emergency diagnosis.

Figure 32 - Number of consultations per patient per month stratified according to age group for all consultations and symptom related consultations leading to a non-emergency and emergency diagnosis



4.8.3.6 Symptom at presentation

Young patients were more likely to present with bleeding (14.0%) than all other age groups except the 50 – 59 year olds. Abdominal pain (21.1%) and ‘other bowel function’ (3.9%) were most common in the youngest age group, while change in bowel habit, anaemia and weight loss increased with age (table 70). No difference was seen with fatigue and an abdominal mass. The under 50 group had the lowest percentage of red-flag (28.5%) and the highest percentage of consultations with only non-specific symptoms (31.5%) at presentation, although only the latter reached significance ($p < 0.005$).

Table 70 - Symptom at presentation after excluding the last 28 days before diagnosis

	Under 50		50 - 59		60 - 69		70 - 79		All		P value
	n	%	n	%	n	%	n	%	n	%	
Red-flag symptom	145 _a	28.5	353 _a	30.2	723 _a	31.5	1085 _a	32.4	2306	31.5	0.231
Only non-specific	160 _a	31.5	285 _b	24.4	549 _b	23.9	812 _b	24.3	1806	24.7	< 0.005
Bleeding	71 _a	14.0	186 _a	15.9	287 _b	12.5	329 _c	9.8	873	11.9	< 0.0005
Anaemia	20 _a	3.9	52 _a	4.5	161 _b	7.0	341 _c	10.2	574	7.8	< 0.0005
Change bowel habit	19 _a	3.7	55 _a	4.7	146 _b	6.4	168 _a	5.0	388	5.3	0.030
Diarrhoea	32 _{a,b}	6.3	64 _{a,b}	5.5	126 _b	5.5	238 _a	7.1	460	6.3	0.054
Abdominal mass	3 _a	0.6	7 _a	0.6	8 _a	0.3	16 _a	0.5	34	0.5	0.726
Abdominal pain	107 _a	21.1	188 _b	16.1	375 _b	16.1	424 _c	12.7	1094	15.0	< 0.0005
Constipation	24 _{a,b}	4.7	38 _b	3.3	76 _b	3.3	203 _a	6.1	341	4.7	< 0.0005
Fatigue	10 _{a,b}	2.0	26 _{a,b}	2.2	48 _b	2.1	105 _a	3.1	189	2.6	0.054
Weight loss	6 _{a,b}	1.2	17 _{a,b}	1.5	31 _b	1.4	79 _a	2.4	133	1.8	0.016
Other bowel function	20 _a	3.9	29 _b	2.5	41 _b	1.8	38 _c	1.1	128	1.7	< 0.0005

Similar letter subscripts indicate no significance difference between groups, with significance set at < 0.05

All age groups were significantly less likely to present to their GP in the year leading up to diagnosis compared with young patients in logistic regression modelling. Young patients were also more likely to present with abdominal pain and ‘other bowel function’ (table 71). Those aged between 70 – 79 were the only group less likely than young patients to present with rectal bleeding (OR 0.66 95% CI 0.50 – 0.87 $p < 0.005$) but were more likely to present with anaemia (OR 1.83 95% CI 1.17 -2.88 $p < 0.05$). There was no difference if the symptoms were grouped into ‘red-flag’ but all age groups were significantly less likely to present with a

non-specific symptom (OR 0.73 CI 0.59 – 0.92, OR 0.68 CI 0.55 – 0.84, OR 0.68 CI 0.55 – 0.83, respectively).

Table 71 - Odds ratio for presentation with a particular symptom

		OR	95% CI	p value
Any symptom	Under 50	ref		
	50 - 59	0.72	0.59 - 0.89	0.002
	60 - 69	0.73	0.60 - 0.88	0.001
	70 - 79	0.73	0.61 - 0.89	0.001
Red- flag	Under 50	ref		
	50 - 59	0.93	0.75 - 1.17	0.548
	60 - 69	1.00	0.81 - 1.23	0.994
	70 - 79	1.02	0.83 - 1.25	0.887
Non-specific	Under 50	ref		
	50 - 59	0.73	0.59 – 0.92	< 0.005
	60 - 69	0.68	0.55 – 0.84	< 0.0005
	70 - 79	0.68	0.55 – 0.83	< 0.0005
Bleeding	Under 50	ref		
	50 - 59	1.07	0.80 – 1.43	0.662
	60 - 69	0.87	0.65 – 1.14	0.309
	70 - 79	0.66	0.50 - 0.87	< 0.005
Anaemia	Under 50	ref		
	50 - 59	1.01	0.61 - 1.69	0.974
	60 - 69	1.38	0.87 - 2.20	0.169
	70 - 79	1.83	1.17 - 2.88	< 0.05
Abdominal pain	Under 50	ref		
	50 - 59	0.76	0.59 - 0.99	0.044
	60 - 69	0.74	0.58 - 0.94	0.013
	70 - 79	0.55	0.43 - 0.70	< 0.0005
Other bowel function	Under 50	ref		
	50 - 59	0.57	0.32 - 0.99	0.047
	60 - 69	0.46	0.27 - 0.77	<0.005
	70 - 79	0.33	0.20 - 0.57	<0.0005
Change in bowel habit	Under 50	ref		
	50 - 59	1.04	0.63 - 1.70	0.890
	60 - 69	1.52	0.96 - 2.39	0.072
	70 - 79	1.35	0.86 - 2.12	0.197
Diarrhoea	Under 50	ref		
	50 – 59	0.76	0.50 – 1.15	0.197
	60 – 69	0.80	0.54 – 1.17	0.251
	70 - 79	1.00	0.69 – 1.46	0.988
Weight loss	Under 50	ref		
	50 – 59	1.21	0.47 – 3.10	0.689
	60 – 69	1.14	0.47 – 2.75	0.776
	70 - 79	1.20	0.85 – 4.62	0.916

Adjusted for site, IMD, Charlson score, gender and emergency presentation. There was no difference noted with change in bowel habit, weight loss and diarrhoea.

4.8.3.7 Symptoms associated with an emergency diagnosis

All age groups (50 – 59: 0.61 CI 0.48 – 0.77 $p < 0.0005$; 60 – 69: 0.62 CI 0.50 – 0.77 $p < 0.0005$; 0.66 CI 0.53 – 0.82 $p < 0.0005$), were less likely to be diagnosed as an emergency compared with the under 50-year-old patients (table 72). Lower deprivation (all except IMD 2 with reference to IMD 1), higher Charlson score (one OR 2.00 95% CI 1.72 – 2.32 $p < 0.0005$, more than one OR 2.07 95% CI 1.72 – 2.49 $p < 0.0005$) and colon cancer (OR 3.45 95% CI 2.95 – 4.03) were associated with an emergency diagnosis.

Table 72 – Patient demographics and tumour site associated with an Emergency Diagnosis

		OR	95% CI	p value
Age	Under 50		ref	
	50 - 59	0.61	0.48 – 0.77	< 0.0005
	60 - 69	0.62	0.50 – 0.77	< 0.0005
	70 - 79	0.66	0.53 – 0.82	< 0.0005
Gender	Male	0.91	0.81 – 1.03	0.126
	Female		ref	
Charlson	Zero		ref	
	1	2.00	1.72 – 2.32	< 0.0005
	≥2	2.07	1.72 – 2.49	< 0.0005
IMD	1		ref	
	2	1.08	0.85 – 1.20	0.934
	3	1.21	1.01 – 1.44	0.036
	4	1.25	1.05 – 1.50	0.015
	5	1.49	1.23 – 1.81	< 0.0005
Site	Colon	3.45	2.95 – 4.03	< 0.0005
	Rectal		ref	

Patients who presented with non-specific symptoms were more likely to be diagnosed as an emergency if they were young compared with all other age groups (50 – 59: 0.51 CI 0.37 – 0.72; 60 – 69: 0.58 CI 0.43 – 0.79; 70 – 79 0.59 0.44 – 0.79 respectively) (table 73). A similar finding was noted for patients presenting with abdominal pain, although no difference was seen if patients presented with rectal bleeding. Compared with young patients, the 60 – 69 year olds (OR 0.52 95% CI 0.30 – 0.93 $p = 0.028$) was the only group less likely to be diagnosed as an emergency if the ‘red-flag’ symptoms were considered together. ‘Other bowel function’ as a presenting symptom was also less likely to lead to an emergency

diagnosis (0.10 CI 0.01 – 0.84) only for the 50 – 59 age group. There were no associations between the other symptoms and an emergency diagnosis when stratified according to age.

Table 73 - Odds ratio (OR) for emergency presentation dependent on the presenting symptom

		OR	95% CI	p value
All patients	Under 50		ref	
	50 - 59	0.61	0.48 – 0.77	< 0.0005
	60 - 69	0.62	0.50 – 0.77	< 0.0005
	70 - 79	0.66	0.53 – 0.82	< 0.0005
Red- flag	Under 50		ref	
	50 - 59	0.54	0.28 – 1.03	0.060
	60 - 69	0.52	0.30 – 0.93	0.028
	70 - 79	0.68	0.39 – 1.18	0.166
No red-flag present	Under 50		ref	
	50 - 59	0.62	0.47 – 0.82	0.001
	60 - 69	0.62	0.49 – 0.80	< 0.0005
	70 - 79	0.63	0.50 – 0.81	< 0.0005
Bleeding	Under 50		ref	
	50 - 59	0.88	0.26 – 3.00	0.840
	60 - 69	0.38	0.11 – 1.31	0.124
	70 - 79	0.68	0.21 – 2.19	0.517
Abdominal pain	Under 50		ref	
	50 - 59	0.61	0.36 – 0.79	0.002
	60 - 69	0.49	0.31 – 0.78	< 0.005
	70 - 79	0.62	0.39 – 0.88	< 0.005
Other bowel function	Under 50		ref	
	50 - 59	0.10	0.01 – 0.84	0.035
	60 - 69	0.38	0.10 – 1.48	0.163
	70 - 79	0.77	0.20 – 3.00	0.700

Adjusted for Charlson score, IMD, gender, site of cancer. No differences were seen with anaemia, change in bowel habit, Diarrhoea and Weight loss.

4.8.3.8 Referral interval

From the initial cohort, 2928 (37.8%) patients were identified with a recording of a relevant symptoms and a referral for cancer, of which 216 (7.4%) were under 50, 467 (15.9%) were 50 – 59, 912 (31.1%) 60 – 69 and 1333 (45.5%) 70 – 79. The median time from presentation to referral was longer for the youngest age group (27 days IQR 1 – 101) compared with those aged 50 – 59 and 60 – 69, with a difference of 6 days (table 74). The time from referral to diagnosis was significantly longer in the youngest age group (59 days IQR 35 – 105), with a difference was 10 days compared to the 60 – 69 year olds. There was no difference from presentation to diagnosis.

Table 74 - The median and interquartile referral, hospital and diagnostic interval in days compared with age

	Under 50	50 - 59	60 - 69	70 - 79	All	p
Presentation to referral	27 (1 – 101)	21.5 (1 – 104)	21 (1 – 91.5)	28 (3 – 117.25)	25 (2 – 105)	0.028
Referral to diagnosis	59 (35 – 105)	46.5 (28 – 85.25)	49 (29 – 83)	47 (28 – 87.25)	49 (28 – 86.75)	0.018
Presentation to diagnosis	108 (60 – 225)	91.5 (54 – 198)	92 (54 – 189)	100 (55 – 216.25)	97 (55 – 206)	0.101

4.8.4 Discussion

4.8.4.1 Summary

The aim of this study was to determine whether young patients experience missed opportunities to identify an underlying CRC in primary care leading to an emergency diagnosis or a delay in referral. I found an emergency diagnosis and a late stage cancer was more common in young patients. Although the percentage of patients with at least one CRC associated presentation did not differ with age, young patients were most likely to present with a non-specific symptom. Compared with older patients also presenting with non-specific symptoms, younger patients were more likely to be diagnosed as an emergency. However, there was no difference in the likelihood of an emergency presentation if the presenting consultation included a red-flag. For young patients with a referral recorded in CPRD, the interval from presentation to referral and from referral to diagnosis was significantly longer, which reflects the higher proportion of non-urgent referrals. My study suggests young patients are experiencing delays in referral and missed opportunities for a non-emergency diagnosis.

4.8.4.2 Relevance to clinical practice

Primary care physicians are aware that cancer is less likely to be the cause of a young patients' symptoms²⁸, explaining the longer interval from presentation to referral in this group, albeit by a median difference of seven days. During this period the GP may organise tests in primary care for an assumed benign condition, only referring if the symptoms persist. My findings demonstrated a relatively high proportion of young patients were non-urgent referrals, whereas those between 60 – 69 years were most commonly referred by the two-week-wait route. This reflects the suspicion of cancer and the current referral guidelines, leading to a median difference of 13 days between referral and diagnosis. However I do not believe the longer intervals are sufficient to account for the advanced stage described for young patients, which instead may be explained by the higher proportion of emergency diagnoses¹⁰⁰¹⁸¹.

The likelihood of an emergency diagnosis did not differ with age if a red-flag symptom was noted at first presentation suggesting the GP considers a cancer diagnosis following these symptoms, irrespective of age . However a non-specific symptom in young patients was more likely to lead to an emergency diagnosis, confirming the belief that GPs are presuming these symptoms are not due to cancer. In addition, five months before an emergency diagnosis there was a relative increase in the consultation rate for young patients, which was not seen in the older age groups, nor in those diagnosed via a non-emergency. This represents a period during which there was a missed opportunity to avoid an emergency diagnosis.

In both the UK and Denmark, GPs play a similar gatekeeper role for access to investigations and secondary care³⁰. Therefore it has been suggested delayed diagnosis in primary care may partly explain the worse survival in these countries compared with their northern and western European neighbours, especially as the majority of patients diagnosed as an emergency have previously presented to their GP^{91,101,182}. As 31% of young patients did not report a red-flag symptom at first presentation, many patients within this age group will remain at risk of missed diagnoses unless there is a better understanding in primary care of how young CRC patients present. This study adds to this growing area of research.

Whether in practice an emergency diagnosis can be prevented in a significant proportion of young patients remains debatable, considering a GP may only see up to two patients per year with one of the four most common cancers²⁸. Also, the reasons why patients are diagnosed as an emergency is complex, with missed opportunities in primary care not the only issue⁸⁸.

The most recent referral guidelines have now included abdominal pain, with the threshold of 40 years of age and weight loss, or 50 years of age with rectal bleeding⁵⁰. This may make the GP more aware of the association of this symptom and cancer, even in younger patients. By simply removing the age criterion we would expect a reduction in emergency diagnoses. However this will significantly reduce the sensitivity of the referral guidelines, thereby

overwhelming secondary care with patients with an underlying benign condition. Alternative methods of detection in primary care need to be considered than simply relying on the type of symptom, especially as the symptom profile in young patients will change as rectal tumours become more common¹⁸³.

4.8.4.3 Comparison to previous studies

This is the first study to describe the most common symptoms in young patients, and the symptoms most likely to lead to an emergency diagnosis in this group.

Similar to previous studies from both the UK and the United States, this study demonstrated the proportion of new diagnoses as well as late stage cancer has increased in young patients over the last 10 years¹¹⁰. Data of patients experiences of primary care in England has shown young patients with colon or rectal cancer are more likely than those aged 65 – 74 to see their GP three or more times before referral, which confirms my findings regarding referral delays in this group¹⁸⁴.

In previous studies, there is inconsistency as to whether a change in bowel habit, diarrhoea or weight loss are associated with an emergency diagnosis, perhaps due to variation in the definition of these symptoms, while rectal bleeding and anaemia seem to be protective¹⁸⁵. The symptoms deemed relevant in this study were agreed by the authors, and have previously been used in other studies to determine intervals from presentation to diagnosis^{52,97,105,168}. Although anaemia may not be a presenting complaint for most patients, it nevertheless should prompt a referral from primary care without the need for an associated symptom.

The interval from presentation to referral in this study was longer than that noted by Lyratzopoulos et al.¹⁶¹ (median 6 days IQR 0 – 29) in their study using the English National Audit of Cancer Diagnosis in Primary Care. Whereas we have calculated the interval from non-specific symptoms in the year before diagnosis, a GP may have not consider this a relevant symptom before submitting the audit data. In addition, the latter is susceptible to

selection bias if the GP felt a delay in primary care may have led to a worse outcome for their patient.

4.8.4.4 Strengths and limitations

By using a large national dataset I was able to describe in detail the difference in presenting symptoms and how young patients are managed in primary care despite presenting with similar symptoms. CPRD is not liable to recall or selection bias as the consultation is recorded at the time of presentation, without the confirmation of the underlying diagnosis. I was also able to determine the difference in the route of referral for non-emergency patients as this information was provided by NCIN using the protocol described by Elliss-Brookes et al.²⁹ By accounting for the socioeconomic deprivation level and Charlson co-morbidity score, and by excluding patients over the age of 79, I was able to remove potential confounders to emergency diagnosis, as noted in previous studies⁵⁵. By using the Generalised Estimating Equation modelling in SPSS, I could adjust for clustering within practices.

Without the availability of the free text from the patients' electronic medical record, CPRD is liable to missing data. If, by definition, all patients diagnosed through the non-emergency pathway would have seen their GP at least once, then we found no significant difference between the age groups in the proportion of patients with at least one consultation recorded in CPRD (Under 50: 65.0% (n = 234); 50 – 59: 58.2% (n = 545), 60 – 69: 59.7% (n = 1084); 70 – 79: 60.0% (n = 1553)). This indicates missing data does not influence a particular age group. The free text may also hold additional symptoms, as only the main complaint is Read-coded. I was unable to ascertain with certainty whether the symptoms recorded were due to the cancer, although this is an issue with any observational study investigating delays.

Stage at diagnosis has often been used as a surrogate measure of quality of care between age groups and between countries. In this study, as in previous studies looking at stage, missing was included as late stage as these two groups are often found to have a

comparable mortality¹⁸¹, to demonstrate young patients were being diagnosed at a later stage. More up to date stage data, which would be expected to have lower missing data, is required to confirm this finding.

I only used one year of data preceding the cancer diagnosis to search for associated symptoms because previous studies have shown medical seeking behaviour changes in this period relative to those with an underlying benign condition. With CPRD it is not possible to include the duration of symptoms before presentation. This interval may have influenced both the urgency of referral and by extension, the stage at diagnosis. The READ coding system alone does not provide sufficient detail to determine this. Although this may have some influence on our findings, two systematic reviews^{31,33}

into the factors that impact patient presentation concluded that the symptoms and co-morbidities had the greatest influence, while age and gender had little or no impact. In our study we have included these possible confounders.

4.8.4.5 Conclusion

Our study concludes that young patients are more likely to present with non-specific symptoms that fall outside the national referral criteria, such as abdominal pain, compared with older patients. Primary care physicians should be made aware of these differences if there is to be a reduction in missed opportunities to prevent emergency diagnoses.

End of Study

CHAPTER FIVE - DISCUSSION

5.1 Summary

The findings from this thesis support the main hypothesis that delays in primary care are associated with a worse outcome for colorectal cancer patients. In addition this thesis improves our understanding of the management of cancer patients in primary care. The findings are summarised below:

- Presentation with non-specific symptoms is associated with missed opportunities for diagnosis in primary care, leading to an emergency diagnosis
- One fifth of colon cancer and one quarter of rectal cancer patients diagnosed as an emergency have previously presented to their GP with a red-flag symptom
- Non-specific symptoms are associated with a late stage disease for both emergency and non-emergency patients
- The type of symptom at presentation impacts the referral interval, but has little effect on the hospital interval
- Delays in referral of over three months for patients presenting with red-flag symptoms are associated with a worse stage and outcome.
- Delays in referral for patients with only non-specific symptoms does not impact survival
- Young patients are more likely to present with non-specific symptoms, but the likelihood of an emergency diagnosis is higher compared to older patients presenting with similar symptoms
- Survival for young patients is better than older patients, despite a worse stage at diagnosis
- Referral by the urgent pathway reduces the referral and hospital interval, but does not improve survival compared to the non-urgent pathway
- CPRD is representative of the national cancer population and is a reliable source to determine delays in primary care

From this thesis I can conclude that the decision to refer by the GP, in terms of speed and the route, is primarily dependent on the symptom at presentation. At every consultation concerning symptoms, the GP has to make the difficult decision whether an underlying cancer has led to the patients presentation. This is a difficult task, considering the very few number of cancer patients a single GP is likely to see per year compared with the larger number of patients with an underlying benign condition. A GP has a limited time per consultation, often less than ten minutes, to make this decision based on the history and clinical examination. Understandably a GP will rely heavily on the NICE referral criteria to guide to patient management. However such guidelines seem to be 'blinding' the GP away from the significant proportion of patients with non-specific symptoms.

Despite the NICE referral guidelines, this thesis has identified a significant proportion of patients who should be identified in primary care and referred to hospital promptly. For example, 21.3% of colon cancer and 25.9% of rectal cancer patients diagnosed as an emergency presented at least once with a red-flag symptom, and so should have avoided the emergency route. In addition, the 31.1% of colon and 24.8% of rectal cancer patients who despite presenting with a red-flag symptom, experienced three or more consultations before referral. The symptoms and patient demographics that were most likely to lead to a delay in referral, were also the factors most associated with an emergency diagnosis - female patients, those either over the age of 70 or under 50, from a lower socioeconomic status and those with a higher Charlson score.

Emergency patients have a worse prognosis than non-emergency patients, as do non-emergency patients with red-flag symptoms referred after three or more consultations or after three months compared with those referred earlier. Those presenting with non-specific symptoms did not demonstrate a worse prognosis despite a three month delay, suggesting such patients are presenting further along the tumour progression. On the contrary, those presenting with red-flag symptoms, and in particular rectal bleeding seem to be presenting before the cancer can progress, and so provide an opportunity for a diagnosis at an earlier stage.

Once a GP has made the decision to refer, the route of referral does not seem to make a difference to the outcome. Referrals taken over three months, or the presenting symptom, have a greater influence. Patients who have been referred by the urgent route, wait 11 days less than those referred by the non-urgent pathway. This alone is insufficient to worsen stage at diagnosis. However despite the accumulative effect of further delays in hospital for non-urgent patients, the outcome and stage at diagnosis between the two routes is non-significant. Although patients referred by the non-urgent pathway seem to be diagnosed at an earlier stage, this is balanced by the higher proportion of missing stage. Patients with missing stage most closely resemble late stage disease terms of survival¹⁸¹. A lower proportion of stage I disease in the TWW group may be explained by the waiting time paradox described in previous studies. As these patients do not experience delays, changing national policy regarding referrals will have little impact on their prognosis. Their aggressive tumour is a non-modifiable factor.

Throughout this thesis, it is apparent young patients defined as those under the age of 50, are most likely to experience delays, whether in terms of emergency diagnoses or referrals. A young patient presenting with colorectal related or non-specific symptoms, is much more likely to have an underlying benign condition, reflecting the increased time taken by GPs to make a decision to refer. In addition to this, this thesis demonstrates a higher proportion of young patients present with non-specific symptoms, further adding difficulty in identifying cancer within this group.

5.2 Implications for clinical practice

As described in the introduction, delays in primary care is one area of improvement identified by the Kings Fund²⁷ and Cancer Research UK to improve the outcome for cancer patients in England and the UK. This thesis has identified a proportion of cancer patients in the UK who despite presenting to their GP with red-flag symptoms, are still diagnosed as an emergency or after a delay in primary care. Improved GP awareness of the symptoms associated with colorectal cancer, and the potential impact of a delayed referral in particular

for those with red-flag symptoms, will reduce delays. Alternative methods of identifying patients with cancer include algorithms which consider patient symptoms¹⁸⁶, demographics and blood results to determine the likelihood of a cancer diagnosis. Access to such programmes may provide the GP an objective evaluation of the chances of cancer on an individual patient basis. However the decision to refer will still be dependent on a national threshold, which is a significant factor in identifying cancer patients.

There have been regional initiatives to allow direct access to endoscopy for patients who fulfil the NICE referral criteria¹⁸⁷. Although this will reduce the hospital interval, it may not have an impact on the referral interval, which remains the significant factor. Perhaps without the need for referrals requiring vetting in secondary care, GPs may be more willing to investigate patients with an endoscopy. As has been shown in various studies^{54,62}, patients from GP practices with a lower internal threshold for referral demonstrate a better outcome.

By the bedside tests in the GP practice have the potential to reduce referral intervals. Introduced in the last five years, faecal occult blood tests, can give an immediate indication as to the likelihood of a cancer diagnosis¹⁸⁸. Proximal cancers, which are less likely to have occult bleeding, are also the cancers most likely to be associated with non-specific symptoms. Therefore current FOBT, although part of the latest NICE referral guidelines⁵⁰, are not sufficiently reliable to be the only reason for referral. New technologies, based on alternative biomarkers, are a possible alternative, but remain untested¹⁸⁹.

The most recent referral guidelines, introduced in 2015⁵⁰, were the first to be based on a PPV, set at 3%. The 2005 guidelines were based on studies⁵⁰, without a PPV set, while the 2000 version was mostly expert opinion. The latest guidelines have now included abdominal pain and weight loss as a criteria for referral. By introducing these symptoms, the proportion of patients who are referred will inevitably increase, and may shorten the referral and diagnostic interval for the 19.7% of patients who complained of abdominal pain in the year before diagnosis. However 100 patients with a combination of these symptoms

would have to be referred for only 3.4 to be diagnosed with cancer. In addition, as shown in this thesis, cancers with these symptoms are likely to be a late stage disease with a worse prognosis, even after excluding patients who were referred more than two weeks from first presentation. Delays in referral for patients with non-specific symptoms are not associated with a worse prognosis. These findings suggests cancers which present with these symptoms will not be diagnosed at an earlier stage simply by a change in referral criteria.

Bowel screening with a flexible endoscopy is an alternative method that will identify cancer patients of at a lower stage before they become symptomatic⁶⁷. However this investigation is currently confined to a patient age where the likelihood of cancer is higher, and only to tumours beyond the splenic flexure, which are less likely to present with non-specific symptoms.

Lowering of the threshold, whether in terms of age or type of symptom, has the potential of both increasing the number of cancer patients, but also overwhelming secondary care services with false negative patients. The most viable solution to reduce late stage diagnosis will be a combination of the methods described, as within the cancer population there is variation in age, symptoms and demographics. Improving identification of cancers with non-specific symptoms may improve outcome of non-emergency patients and also reduce the proportion of emergency cases, with a resultant improvement in prognosis.

Other areas of potential improvement outside the scope of this thesis include patient awareness and post diagnosis management. In study III regarding delays in adjuvant chemotherapy following colorectal cancer treatment, has shown the latter can be explored in CPRD. The awareness of the patient to a potential cancer diagnosis will determine the promptness of presentation, and will conceivably influence outcome.

5.3 Comparison to previous studies

The findings of this thesis are consistent with previous findings. Patients with red-flag symptoms, such as rectal bleeding are likely to present with earlier stage disease and

demonstrate better prognosis¹⁷³. Emergency patients demonstrate a later stage and worse prognosis⁸⁸. Young patients are diagnosed at a later stage compared with those age 60 – 69, but nevertheless demonstrate a better prognosis.

This is the first study to specifically analyse the referral interval, and its influence on patient outcome, while accounting for patient co-variables that have been shown to influence delays and outcome. Previous studies investigating delays in primary care have analysed the entire interval from first presentation to the GP to diagnosis. This study has shown the hospital interval, defined as the interval from referral to diagnosis, does not vary sufficiently between patients referred within two weeks and beyond three months to explain the worse outcome in the latter. This thesis is therefore the first to suggest a limit to how long a GP should wait before referral.

Studies by Topping et al.^{78,79} have suggested the interval from first presentation to diagnosis should not be considered a linear relationship to prognosis, but rather U-shaped. As such patients with red-flag symptoms and a relatively short diagnostic interval demonstrate a worse prognosis, as do patients with a significantly prolonged interval. Such a relationship was not seen in patients with non-specific symptoms. These findings have been replicated with data from both Denmark and the UK, using different sources^{79,80}. This thesis goes some way in confirming these findings, as patients referred over three months with red-flag symptoms demonstrated a worse prognosis, a finding not seen for patients with non-specific symptoms. This thesis adds to the Topping et al. ^{78–80} studies by suggesting the referral interval, rather than the whole diagnostic interval, determines outcome. This narrows the target to improve outcome on a national basis. For future research, the referral interval could be analysed in a similar manner to the Topping et al. ^{78,79} study to determine if a U-shaped relationship is found.

5.4 Limitations

I believe the aims set out at the beginning of each chapter have been met. However the findings of the thesis need to be considered with the following limitations.

5.4.1 Missing data

As stated in the introduction, CPRD is considered robust in terms of diagnosis¹²⁷. However in only a handful of studies has it been used as a database to determine the number of consultations, or the length of time between clinical events. Instead its main purpose has been in epidemiological studies which rely on the presence or absence of a diagnosis, rather than the level of detail that is required for this thesis¹²¹. This outlines the disadvantages of routinely collected national data. Although CPRD is relatively cheap in terms of resources compared to bespoke data, the cost and time to collect for the latter makes it prohibitive. Throughout this thesis the accuracy of CPRD has been tested, which is outlined here.

During a consultation, the GP is required to enter a Read code before the clinical encounter can be logged. The detail of this Read code varies, and determines the accuracy of the data used in this thesis. In a study by Price et al.¹²⁴, the authors analysed the free text and the Read code in CPRD for haematuria, jaundice and abdominal pain in bladder and pancreatic cancer cases. These were chosen as the former two are 'alarm' symptoms for the two cancers, and the latter is considered a 'low-risk' symptom for both cancers. This was compared with patients without cancer, primarily to determine the effect on the PPV of data hidden in the text. They found alarm symptoms were more likely to be recorded as Read codes and less likely to be hidden in the free text than non-specific symptoms, thereby falsely inflating the PPV. They concluded GPs record the symptoms as a Read code for the symptom most likely to be associated with cancer, which have been defined as red-flag in this thesis.

There are possible effects of this on the findings of this thesis. At the first consultation, I found the vast majority of patients only had one symptom recorded. Also those with red-flag symptoms were less likely to have an additional symptom. In contrast previous studies have shown a patient may present with rectal bleeding and a change in bowel habit towards diarrhoea⁴¹. For such a patient, ideally the patients would code each symptom separately, thereby providing three medcodes for one consultation. Based on the study by Price et al¹²⁴. and the PPV of symptoms, I organised patients according to red-flag and non-specific

symptoms. It is likely that the patients defined as having red-flag symptoms also presented with non-specific symptoms which were recorded in the free text. Conversely patients with non-specific symptoms were unlikely to have an associated red-flag symptom during the consultation. Without access to the free text, it's not possible to determine the proportion of patients with truly more than one symptom. Therefore the worse outcome associated with non-specific symptoms is still likely to be true.

Alternatively some missing data may be explained by the medcodes chosen to represent the symptoms, which were based on previous studies. Instead of medcodes describing symptoms, consultations can be logged with non-specific Read terms such as '19...00', defined as 'Gastrointestinal symptoms', with additional information regarding the actual symptom as the free text. This has been recorded 34235 times within CPRD. As previously found, such consultation may be more likely to involve non-specific symptoms. Therefore the higher proportion of missing consultation data associated with emergency patients or those with later stage, may be a result of non-specific consultations, thereby reinforcing these associations.

Practices, and GPs within those practices, are likely to repeat the same pattern of Read code usage. Throughout this thesis I have used the Generalised Estimating Equation module in an attempt to account for clustering within practices.

As described in the narrative review regarding CPRD, there are various methods to determine the accuracy of the data used. For a handful of cases the patients notes can be reviewed to determine whether consultations have been missed. This has been carried out for diagnoses, but not for individual symptom related consultations. Alternatively, the findings from CPRD can be compared to studies that have used patient notes for a similar purpose. For the purpose of this comparison, I have included data in the year leading up to diagnosis, and in particular for rectal bleeding, anaemia and abdominal pain.

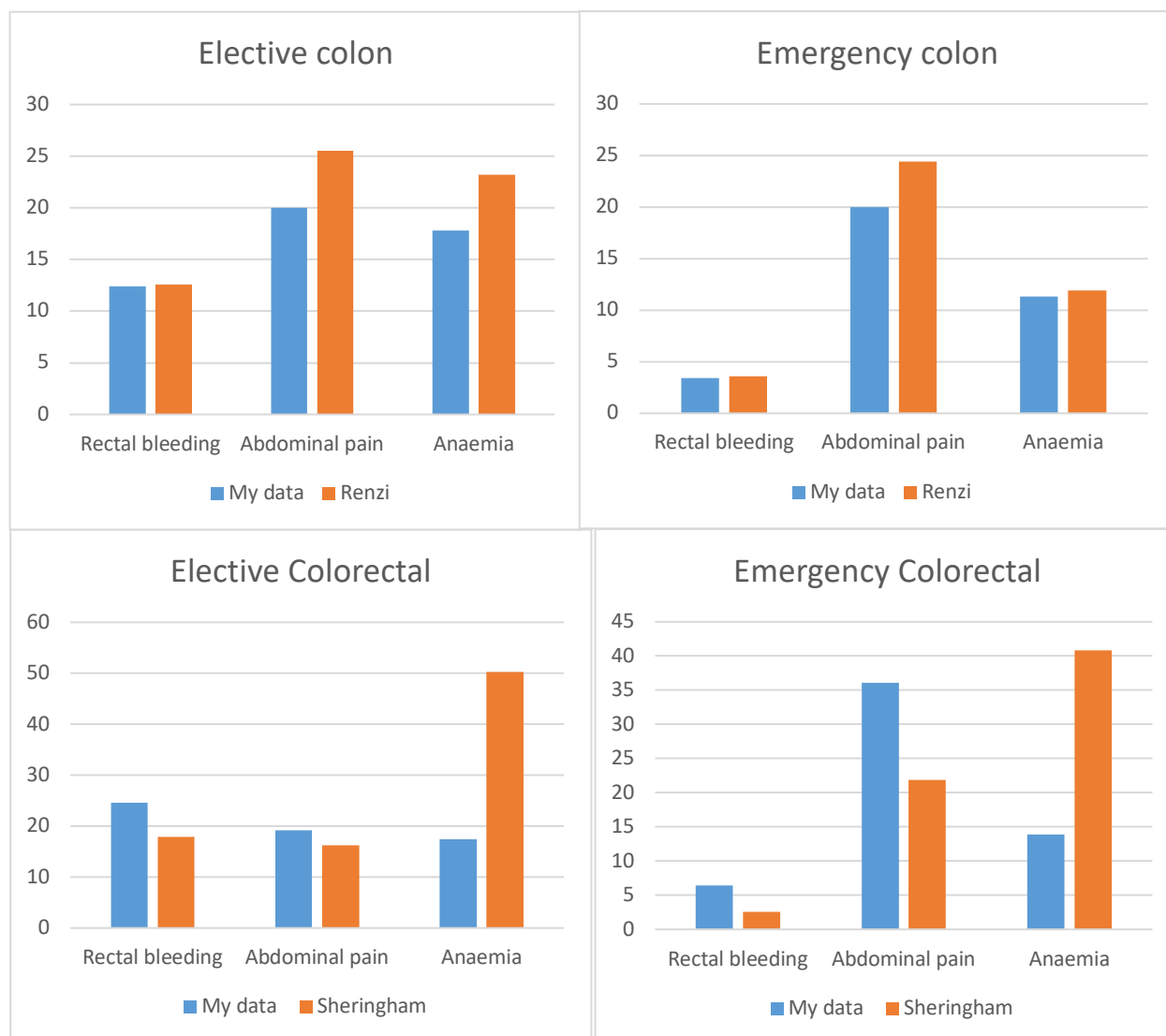


Figure 33 - Comparison of percentage of patients with rectal bleeding or anaemia between my thesis and data from Sheringham et al.⁹⁴

Compared to the study by Sheringham et al.⁹⁷ (figure 33), which used retrospective analysis of patient notes, the percentage of patients with abdominal pain or rectal bleeding was slightly higher in my data for both elective and emergency colorectal cancer diagnoses. However the percentage of patients with anaemia was considerably higher in the Sheringham et al.⁹⁷ data. The difference may be explained by a combination of the source data used, or the definition of these symptoms and anaemia. For example anaemia in my study was based on the medcodes rather than the blood results. Secondly it is not clear from the Sheringham et al.⁹⁷ paper the definitions used to determine the symptoms. Comparison with Renzi et al.¹⁰⁵ which also used CPRD, demonstrates the proportion of

patients with particular symptoms is more similar, suggesting a difference in the data source, rather than my methodology.

All of the 10661 patients who were diagnosed as a non-emergency should have had a symptom recorded in the year leading up to cancer diagnosis. However there seems to be a tendency for certain patient groups to have missing records, namely male patients, those with a Charlson score, higher socio-economic deprivation and later stage. It may be argued that these same characteristics are associated with an emergency diagnosis, and so are perhaps due to non-emergency patients incorrectly being classified as emergency cases. However sensitivity analysis by inclusion of only patients diagnosed after 2006 inclusive, for whom a route of diagnosis is available, reveals similar results. It is feasible that patients with a higher Charlson score are consulting their GP on a regular basis for other reasons, leading to their symptoms being recorded within the free text rather than as a Read codes.

It can be assumed that all the patients diagnosed as a non-emergency should have a referral to secondary care. However there was a significant amount of missing referral. The Read terms for referral were either based on a symptom, or the specialty who was receiving the referral. Again if a patient was being referred because of a non-specific symptom, this could have been missed. Also vague referrals, such as those to 'Medical Speciality', were not included as it was unclear if this was due to the presenting symptom.

5.4.2 Observational rather than randomised controlled study

This finding of this thesis is based on observational studies using data that has been analysed retrospectively. As such it is not possible to make conclusions regarding causation, but rather association. However the use of randomised control trials regarding the route of diagnosis, or delay before referral, would not be acceptable ethically. Therefore observational studies as described in this thesis are the only reliable method to investigate delays in primary care.

5.4.3 Symptoms included in analysis

As described above, the accuracy of symptom recording as a Read term is dependent on the perceived association with cancer. However which symptoms should have been included? The symptoms included in this thesis are based on previous studies, both CPRD and non-CPRD, the NICE guidelines and our own clinical experience. As noted in study IV. the referral interval differs between change in bowel habit, diarrhoea, constipation and 'other bowel function', indicating the GP is managing the patients differently based on these symptoms. It is unlikely the majority of patients with anaemia or an abdominal mass were aware of these symptoms and therefore presented complaining of either. However both these descriptions are a reason to refer to secondary care, present in the NICE guidelines, which influence the management of patients in primary care.

5.4.4 Stage data

After cleaning the stage data, the proportion of each stage and missing is similar to national data. Missing stage data has been imputed based on patient demographics⁸¹. This was an option I considered, but was concerned there would be an inherent bias introduced as the level of missing stage data differs according to patient demographics and the type of presenting symptom. Therefore imputed stage data as a dependent variable in regression analysis would be inappropriate.

5.5 Further research

The latest NICE guidelines were introduced in 2015, after the end of data coverage used in this study. An up to date dataset may indicate a reduction in primary care delays than has been demonstrated in this thesis.

Outcomes in Denmark, another country where the GP plays a gatekeeper role, may also be associated with delays in primary care⁵³. If the findings from Denmark correlate to the findings in this thesis, it adds further emphasis about the influence of primary care delays on

outcome. Torring et al. ^{78,79} have made comparisons between these two countries to find similar outcome regarding the interval from first symptom to diagnosis. In addition, comparisons with the better performing countries in Europe would also be informative.

To overcome the missing symptom data, it would be necessary to review individual patient notes. This would improve the capture of combination of symptoms, which is missing from CPRD. Although it would not be possible to include as many patients as were available in CPRD due to time and monetary restraints, individual patient notes review, with better stage data, could verify the findings of this study.

5.5 Conclusion

Cancer patients with non-specific symptoms are more likely to be missed in primary care, leading to an emergency diagnosis, in particular young patients. Patients who present with red-flag symptoms and experience a delay of over three months before referral, demonstrate a worse stage and prognosis compared to patients referred within two weeks. Despite presenting with red-flag symptoms, a significant proportion of patients still experience emergency diagnosis or delays in referral. This thesis provides a description of the current management of patients in primary care, and areas of improvement if the prognosis of cancer patients is to improve.

References

1. Weller D, Vedsted P, Rubin G, et al. The Aarhus statement: improving design and reporting of studies on early cancer diagnosis. *Br J Cancer*. 2012;106(7):1262-1267. doi:10.1038/bjc.2012.68
2. Lyratzopoulos G, Wardle J, Rubin G. Rethinking diagnostic delay in cancer: how difficult is the diagnosis? *BMJ*. 2014;349(606):g7400. doi:10.1136/bmj.g7400
3. Thorne K, Hutchings HA, Elwyn G. The effects of the Two-Week Rule on NHS colorectal cancer diagnostic services: a systematic literature review. *BMC Health Serv Res*. 2006;6:43. doi:10.1186/1472-6963-6-43
4. Williams T, van Staa T, Puri S, Eaton S. Recent advances in the utility and use of the General Practice Research Database as an example of a UK Primary Care Data resource. *Ther Adv Drug Saf*. 2012;3:89-99. doi:10.1177/2042098611435911
5. Department of Health. The NHS Cancer Plan. *Dep Heal*. 2000;(September):1-98. http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4014513.pdf.
6. Santi M, Capocaccia R, Verdecchia A, et al. Comparisons of colon-cancer survival among European countries: The EURO CARE study. *Int J Cancer*. 1995;63(1):43-48. doi:10.1002/ijc.2910630109
7. Abdel-Rahman M, Stockton D, Rachet B, Hakulinen T, Coleman MP. What if cancer survival in Britain were the same as in Europe: how many deaths are avoidable? *Br J Cancer*. 2009;101 Suppl:S115-24. doi:10.1038/sj.bjc.6605401
8. Arnold M, Rutherford MJ, Bardot A, et al. Progress in cancer survival, mortality, and incidence in seven high-income countries 1995–2014 (ICBP SURVMARK-2): a population-based study. *Lancet Oncol*. 2019;20(11):1493-1505. doi:10.1016/S1470-2045(19)30456-5
9. Bray F, Parkin DM. Evaluation of data quality in the cancer registry: Principles and methods. Part I: Comparability, validity and timeliness. *Eur J Cancer*. 2009;45(5):747-755. doi:10.1016/j.ejca.2008.11.032
10. Parkin DM, Bray F. Evaluation of data quality in the cancer registry: Principles and methods Part II. Completeness. *Eur J Cancer*. 2009;45(5):756-764. doi:10.1016/j.ejca.2008.11.033
11. De Angelis R, Sant M, Coleman MP, et al. Cancer survival in Europe 1999-2007 by country and age: results of EURO CARE-5-a population-based study. *Lancet Oncol*. 2013;2045(13):1-12. doi:10.1016/S1470-2045(13)70546-1
12. Coleman MP, Forman D, Bryant H, et al. Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995-2007 (the international cancer benchmarking partnership): An analysis of population-based cancer registry data. *Lancet*. 2011;377(9760):127-138. doi:10.1016/S0140-6736(10)62231-3
13. Coleman MP, Quaresma M, Berrino F, et al. Cancer Survival in five continents: a worldwide population-based study (CONCORD). *Lancet Oncol*. 2008;9:731-756. doi:10.1016/S1470-2045(08)70179-7
14. Brown S, Castelli M, Hunter DJ, et al. Social Science & Medicine How might healthcare systems in fl uence speed of cancer diagnosis : A narrative review. *Soc Sci Med*. 2014;116:56-63. doi:10.1016/j.socscimed.2014.06.030
15. National Cancer Intelligence. Colorectal Cancer Survival by Stage. *Color Cancer Surviv by Stage*. 2012;(June). Available%5Cnfrom http://www.ncin.org.uk/publications/data_brie%1Dngs/colorectal_%5Cncancer_survival_by_stage.aspx.
16. McPhail S, Johnson S, Greenberg D, Peake M, Rous B. Stage at diagnosis and early mortality from cancer in England. *Br J Cancer*. 2015;112(s1):S108-S115. doi:10.1038/bjc.2015.49
17. Camilloni L, Ferroni E, Cendales BJ, et al. Methods to increase participation in organised screening programs: a systematic review. *BMC Public Health*. 2013;13(1):464. doi:10.1186/1471-2458-13-464
18. Gatta G, Capocaccia R, Sant M, et al. Understanding variations in survival for colorectal

- cancer in Europe: a EURO CARE high resolution study. *Gut*. 2000;47(4):533-538.
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1728079&tool=pmcentrez&render type=abstract>. Accessed June 1, 2015.
19. Ciccolallo L, Capocaccia R, Coleman MP, et al. Survival differences between European and US patients with colorectal cancer: role of stage at diagnosis and surgery. *Gut*. 2005;54(2):268-273. doi:10.1136/gut.2004.044214
 20. Woodman CB, Gibbs A, Scott N, Haboubi NY, Collins S. Are differences in stage at presentation a credible explanation for reported differences in the survival of patients with colorectal cancer in Europe? *Br J Cancer*. 2001;85(6):787-790. doi:10.1054/bjoc.2001.1958
 21. Maringe C, Walters S, Rachet B, et al. Stage at diagnosis and colorectal cancer survival in six high-income countries: a population-based study of patients diagnosed during 2000-2007. *Acta Oncol*. 2013;52:919-932. doi:10.3109/0284186X.2013.764008
 22. Thomson CS, Forman D. Cancer survival in England and the influence of early diagnosis: what can we learn from recent EURO CARE results? *Br J Cancer*. 2009;101(S2):S102-S109. doi:10.1038/sj.bjc.6605399
 23. Sant M, Allemani C, Santaquilani M, Knijn A, Marchesi F, Capocaccia R. EURO CARE-4. Survival of cancer patients diagnosed in 1995-1999. Results and commentary. *Eur J Cancer*. 2009;45(6):931-991. doi:10.1016/j.ejca.2008.11.018
 24. Møller H, Linklater KM, Robinson D. A visual summary of the EURO CARE-4 results: a UK perspective. *Br J Cancer*. 2009;101 Suppl(SUPPL. 2):S110-4. doi:10.1038/sj.bjc.6605400
 25. Index of cancer survival for Clinical Commissioning Groups in England - Office for National Statistics.
<https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/indexofcancersurvivalforclinicalcommissioninggroupsinengland/adultsdiagnosed2000to2015andfollowedupto2016>. Accessed December 9, 2017.
 26. Cancer survival in England - adults diagnosed - Office for National Statistics 2013.
<https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/cancersurvivalratescancersurvivalinenglandadultsdiagnosed>. Accessed December 9, 2017.
 27. Foot C, Harrison T. How to improve cancer survival Explaining England's relatively poor rates. *Kings Fund*. 2011;(June):1-32.
 28. Hamilton W. Five misconceptions in cancer diagnosis. *Br J Gen Pract*. 2009;59(563):441-447. doi:10.3399/bjgp09X420860
 29. Elliss-Brookes L, McPhail S, Ives A, et al. Routes to diagnosis for cancer - determining the patient journey using multiple routine data sets. *Br J Cancer*. 2012;107(8):1220-1226. doi:10.1038/bjc.2012.408
 30. Olesen F, Hansen RP, Vedsted P. Delay in diagnosis: the experience in Denmark. *Br J Cancer*. 2009;101 Suppl(S2):S5-S8. doi:10.1038/sj.bjc.6605383
 31. Macdonald S, Macleod U, Campbell NC, Weller D, Mitchell E. Systematic review of factors influencing patient and practitioner delay in diagnosis of upper gastrointestinal cancer. *Br J Cancer*. 2006;94(April):1272-1280. doi:10.1038/sj.bjc.6603089
 32. Mitchell E, Macdonald S, Campbell NC, Weller D, Macleod U. Influences on pre-hospital delay in the diagnosis of colorectal cancer: a systematic review. *Br J Cancer*. 2008;98(December 2007):60-70. doi:10.1038/sj.bjc.6604096
 33. Macleod U, Mitchell ED, Burgess C, Macdonald S, Ramirez a J. Risk factors for delayed presentation and referral of symptomatic cancer: evidence for common cancers. *Br J Cancer*. 2009;101 Suppl(S2):S92-S101. doi:10.1038/sj.bjc.6605398
 34. Forbes LJL, Warburton F, Richards MA, Ramirez AJ. Risk factors for delay in symptomatic presentation: a survey of cancer patients. *Br J Cancer*. 2014;111(3):581-588. doi:10.1038/bjc.2014.304
 35. Niksic M, Rachet B, Warburton FG, Wardle J, Ramirez AJ, Forbes LJL. Cancer symptom awareness and barriers to symptomatic presentation in England—are we clear on cancer? *Br J*

- Cancer*. 2015;113(3):533-542. doi:10.1038/bjc.2015.164
36. Walter FM, Emery JD, Mendonca S, et al. Symptoms and patient factors associated with longer time to diagnosis for colorectal cancer: results from a prospective cohort study. *Br J Cancer*. 2016;115(5):533-541. doi:10.1038/bjc.2016.221
37. Ramirez AJ, Westcombe AM, Burgess CC, Sutton S, Littlejohns P, Richards MA. Factors predicting delayed presentation of symptomatic breast cancer: A systematic review. *Lancet*. 1999;353(9159):1127-1131. doi:10.1016/S0140-6736(99)02142-X
38. Delaney BC. Why do dyspeptic patients over the age of 50 consult their general practitioner? A qualitative investigation of health beliefs relating to dyspepsia. *Br J Gen Pract*. 1998;48(433):1481-1485.
39. Lyratzopoulos G, Abel GA, McPhail S, Neal RD, Rubin P. 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(3):686-690. doi:10.1038/Bjc.2013.1
40. Lyratzopoulos G, Neal RD, Barbiere JM, Rubin GP, Abel GA. Variation in number of general practitioner consultations before hospital referral for cancer: findings from the 2010 National Cancer Patient Experience Survey in England. *Lancet Oncol*. 2012;13(4):353-365. doi:10.1016/S1470-2045(12)70041-4
41. 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(S2):S80-S86. doi:10.1038/sj.bjc.6605396
42. Sedighi I. Interpretation of Diagnostic Tests: Likelihood Ratio vs. Predictive Value. *Iran J Pediatr*. 2013;23(6):717. <http://www.ncbi.nlm.nih.gov/pubmed/24910762>. Accessed August 18, 2017.
43. Hamilton W, Round A, Sharp D, Peters TJ. Clinical features of colorectal cancer before diagnosis: a population-based case-control study. *Br J Cancer*. 2005;93(4):399-405. doi:10.1038/sj.bjc.6602714
44. Hippisley-Cox J, Coupland C. Identifying patients with suspected colorectal cancer in primary care: derivation and validation of an algorithm. *Br J Gen Pract*. 2012;62(594):e29-37. doi:10.3399/bjgp12X616346
45. Panzuto F, Chiriatti A, Bevilacqua S, et al. Symptom-based approach to colorectal cancer: survey of primary care physicians in Italy. *Dig Liver Dis*. 2003;35(12):869-875. <http://www.ncbi.nlm.nih.gov/pubmed/14703882>.
46. Ford AC, O Veldhuyzen van Zanten SJ, Rodgers CC, et al. Diagnostic utility of alarm features for colorectal cancer: systematic review and meta-analysis. 2008. doi:10.1136/gut.2008.159723
47. Astin M, Griffin T, Neal RD, Rose P, Hamilton W. The diagnostic value of symptoms for colorectal cancer in primary care : *Br J Gen Pract*. 2011;(May):231-243. doi:10.3399/bjgp11X572427.Conclusion
48. Barrett J, Jiwa M, Rose P, Hamilton W. Pathways to the diagnosis of colorectal cancer: an observational study in three UK cities. *Fam Pract*. 2006;23(1):15-19. doi:10.1093/fampra/emi093
49. NICE. *Referral Guidelines for Suspected Cancer*.; 2005.
50. National Initiative for Health and Care Excellence. NICE guidelines for suspected cancer. 2015;(June). <https://www.nice.org.uk/guidance/ng12/>.
51. Department of Health. Referral guidelines for suspected cancer. *Ref Guidel suspected cancer*. 2000;(June 2000):1-98. doi:10.1016/S0140-6736(10)61376-1
52. Neal RD, Din NU, Hamilton W, et al. Comparison of cancer diagnostic intervals before and after implementation of NICE guidelines: analysis of data from the UK General Practice Research Database. *Br J Cancer*. 2014;110(3):584-592. doi:10.1038/bjc.2013.791
53. Vedsted P, Olesen F. A differentiated approach to referrals from general practice to support early cancer diagnosis – the Danish three-legged strategy. *Br J Cancer*. 2015;(March):1-5. doi:10.1038/bjc.2015.44

54. Shawihdi M, Thompson E, Kapoor N, et al. Variation in gastroscopy rate in English general practice and outcome for oesophagogastric cancer: retrospective analysis of Hospital Episode Statistics. *Gut*. 2014;63(2):250-261. doi:10.1136/gutjnl-2012-304202
55. Bottle a, Tsang C, Parsons C, Majeed a, Soljak M, Aylin P. Association between patient and general practice characteristics and unplanned first-time admissions for cancer: observational study. *Br J Cancer*. 2012;107(8):1213-1219. doi:10.1038/bjc.2012.320
56. Arbmán G, Nilsson E, Störgren-Fordell V, Sjö Dahl R. A short diagnostic delay is more important for rectal cancer than for colonic cancer. *Eur J Surg*. 1996;162(11):899-904. <http://www.ncbi.nlm.nih.gov/pubmed/8956960>. Accessed June 2, 2017.
57. Mariscal M, Llorca J, Prieto-Salceda D, Palma S, Delgado-Rodríguez M. Determinants of the interval between diagnosis and treatment in patients with digestive tract cancer. *Oncol Rep*. March 2003. doi:10.3892/or.10.2.463
58. MacArthur C, Smith a. Factors associated with speed of diagnosis, referral, and treatment in colorectal cancer. *J Epidemiol Community Health*. 1984;38(2):122-126. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1052333&tool=pmcentrez&render type=abstract>.
59. 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(17):2467-2478. doi:10.1016/j.ejca.2007.08.023
60. 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(4):510-521. doi:10.1016/j.ejca.2008.01.011
61. Cerdán-Santacruz C, Cano-Valderrama O, Cárdenas-Crespo S, et al. Colorectal cancer and its delayed diagnosis: Have we improved in the past 25 years? *Rev Esp Enfermedades Dig*. 2011;103(9):458-463. <http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L362762149%5Cnhttp://www.reed.es/downloadContenido.php?idContenido=2021&idioma=en%5Cnhttp://sfx.library.uu.nl/utrecht?sid=EMBASE&issn=11300108&id=doi:&atitle=Colorectal+cancer+and+it>.
62. 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? *Color Dis*. 2011;13(11):1242-1248. doi:10.1111/j.1463-1318.2010.02438.x
63. Stapley S, Peters TJ, Sharp D, Hamilton W. The mortality of colorectal cancer in relation to the initial symptom at presentation to primary care and to the duration of symptoms: a cohort study using medical records. *Br J Cancer*. 2006;95(10):1321-1325. doi:10.1038/sj.bjc.6603439
64. Stapley S, Peters TJ, Sharp D, Hamilton W. The mortality of colorectal cancer in relation to the initial symptom at presentation to primary care and to the duration of symptoms: a cohort study using medical records. *Br J Cancer*. 2006;95:1321-1325. doi:10.1038/sj.bjc.6603439
65. Deng SX, An W, Gao J, et al. Factors influencing diagnosis of colorectal cancer: A hospital-based survey in china. *J Dig Dis*. 2012;13(10):517-524. doi:10.1111/j.1751-2980.2012.00626.x
66. Eccersley J, Wilson E, Makris A, Novell R. Referral guidelines for colorectal cancer--do they work? *Ann R Coll Surg Engl*. 2004;86:402. doi:10.1308/1478708041782238
67. Flashman K, O'Leary DP, Senapati a, Thompson MR. The Department of Health's "two week standard" for bowel cancer: is it working? *Gut*. 2004;53(July 2000):387-391. doi:10.1136/gut.2003.020503
68. Neal RD, Allgar VL, Ali N, et al. Stage, survival and delays in lung, colorectal, prostate and ovarian cancer: comparison between diagnostic routes. *Br J Gen Pract*. 2007;57(536):212-219. <http://www.ncbi.nlm.nih.gov/pubmed/17359608>. Accessed August 15, 2016.
69. Møller H, Gildea C, Meechan D, Rubin G, Round T, Vedsted P. Use of the English urgent referral pathway for suspected cancer and mortality in patients with cancer: cohort study. *BMJ*. 2015;351(oct13_4):h5102. doi:10.1136/bmj.h5102

70. Maclean R, Jeffreys M, Ives A, Jones T, Verne J, Ben-Shlomo Y. Primary care characteristics and stage of cancer at diagnosis using data from the national cancer registration service, quality outcomes framework and general practice information. *BMC Cancer*. 2015;15(1). doi:10.1186/s12885-015-1497-1
71. 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. *Color Dis*. 2012;14(7):848-853. doi:10.1111/j.1463-1318.2011.02829.x
72. 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. *Color Dis*. 2012;14(7):e375-e377. doi:10.1111/j.1463-1318.2012.02961.x
73. Public Health England. National Cancer Intelligence Network. Routes to diagnosis 2006–2013 workbook. <http://www.ncin.org.uk/view?rid=3053>.
74. Wattacheril J, Kramer JR, Richardson P, et al. Lagtimes in diagnosis and treatment of colorectal cancer: determinants and association with cancer stage and survival. *Aliment Pharmacol Ther*. 2008;28:1166-1174. doi:10.1111/j.1365-2036.2008.03826.x
75. 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(5):961-977. doi:10.1007/s10552-013-0172-6
76. Singh H, Shu E, Demers A, Bernstein CN, Griffith J, Fradette K. Trends in time to diagnosis of colon cancer and impact on clinical outcomes. *Can J Gastroenterol*. 2012;26(12):877-880. <http://www.ncbi.nlm.nih.gov/pubmed/23248786>.
77. Stapley S, Peters TJ, Sharp D, Hamilton W. The mortality of colorectal cancer in relation to the initial symptom at presentation to primary care and to the duration of symptoms: a cohort study using medical records. *Br J Cancer*. 2006;95(10):1321-1325. doi:10.1038/sj.bjc.6603439
78. 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(6):934-940. doi:10.1038/bjc.2011.60
79. 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(6):669-678. doi:10.1016/j.jclinepi.2011.12.006
80. Tørring ML, Murchie P, Hamilton W, et al. Evidence of advanced stage colorectal cancer with longer diagnostic intervals: a pooled analysis of seven primary care cohorts comprising 11 720 patients in five countries. *Br J Cancer*. 2017;(August):1-10. doi:10.1038/bjc.2017.236
81. 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(5):e0126608. doi:10.1371/journal.pone.0126608
82. Murchie P, Raja EA, Brewster DH, 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(3):461-469. doi:10.1038/bjc.2014.352
83. Gort M, Otter R, Plukker JTM, Broekhuis M, Klazinga NS. Actionable indicators for short and long term outcomes in rectal cancer. *Eur J Cancer*. 2010;46(10):1808-1814. doi:10.1016/j.ejca.2010.02.044
84. Terhaar sive Droste JS, Oort FA, van der Hulst RWM, 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. doi:10.1186/1471-2407-10-332
85. Yun YH, Kim YA, Min YH, et al. The influence of hospital volume and surgical treatment delay on long-term survival after cancer surgery. *Ann Oncol*. 2012;23:2731-2737. doi:10.1093/annonc/mds101
86. Roland CL, Schwarz RE, Tong L, et al. Surgical Outcome Research Is timing to delivery of treatment a reliable measure of quality of care for patients with colorectal adenocarcinoma? 2013. doi:10.1016/j.surg.2013.04.049

87. NCIN. *Routes to Diagnosis Workbook 2006 - 2014*.; 2016.
http://www.ncin.org.uk/publications/routes_to_diagnosis.
88. Zhou Y, Abel GA, Hamilton W, et al. Diagnosis of cancer as an emergency: a critical review of current evidence. *Nat Rev Clin Oncol*. 2016;advance on.
<http://dx.doi.org/10.1038/nrclinonc.2016.155>.
89. Cancer Patient Experience Survey 2014 National Report. <https://www.quality-health.co.uk/resources/surveys/national-cancer-experience-survey/>. Published 2014.
90. McArdle CS, Hole DJ. Emergency presentation of colorectal cancer is associated with poor 5-year survival. *Br J Surg*. 2004;91(December 1994):605-609. doi:10.1002/bjs.4456
91. Cleary J, Peters TJ, Sharp D, Hamilton W. Clinical features of colorectal cancer before emergency presentation: A population-based case - Control study. *Fam Pract*. 2007;24:3-6. doi:10.1093/fampra/cml059
92. Comber H, Sharp L, Cancela MDC, Haase T, Johnson H, Pratschke J. Causes and outcomes of emergency presentation of rectal cancer. doi:10.1002/ijc.30149
93. Lyratzopoulos G, Saunders CL, Abel GA. Are emergency diagnoses of cancer avoidable ? A proposed taxonomy to motivate study design and support service improvement. *Futur Oncol*. 2014;10:10-14. doi:10.2217/FON.14.80
94. Chou C-L, Chang S-C, Lin T-C, et al. Differences in clinicopathological characteristics of colorectal cancer between younger and elderly patients: an analysis of 322 patients from a single institution. *AJS*. 2011;202:574-582. doi:10.1016/j.amjsurg.2010.10.014
95. Pruitt SL, Davidson NO, Gupta S, Yan Y, Schootman M. Missed opportunities : racial and neighborhood socioeconomic disparities in emergency colorectal cancer diagnosis and surgery. 2014:1-10.
96. Gunnarsson H, Ekholm A, Olsson LI. Emergency presentation and socioeconomic status in colon cancer. *Eur J Surg Oncol*. 2013;39(8):831-836. doi:10.1016/j.ejso.2013.04.004
97. Sheringham JR, Georgiou T, Chitnis X a, Bardsley M. Comparing primary and secondary health-care use between diagnostic routes before a colorectal cancer diagnosis: cohort study using linked data. *Br J Cancer*. 2014;111(8):1490-1499. doi:10.1038/bjc.2014.424
98. NCIN. Routes to Diagnosis : Exploring Emergency Presentations NCIN Data Briefing. 2013.
99. NCIN. *Routes to Diagnosis Workbook 2006 - 2013*.; 2015.
http://www.ncin.org.uk/publications/routes_to_diagnosis.
100. McPhail S, Elliss-Brookes L, Shelton J, et al. Emergency presentation of cancer and short-term mortality. *Br J Cancer*. 2013;109(8):2027-2034. doi:10.1038/bjc.2013.569
101. Abel GA, Mendonca SC, McPhail S, Zhou Y, Elliss-Brookes L, Lyratzopoulos G. Emergency diagnosis of cancer and previous general practice consultations: insights from linked patient survey data. *Br J Gen Pract*. 2017:1-11. doi:10.3399/bjgp17X690869
102. Tsang C, Bottle A, Majeed A, Aylin P. Cancer diagnosed by emergency admission in England: an observational study using the general practice research database. *BMC Health Serv Res*. 2013;13(1):308. doi:10.1186/1472-6963-13-308
103. Sikka V, Ornato JP. Cancer diagnosis and outcomes in Michigan EDs vs other settings. *Am J Emerg Med*. 2012;30(2):283-292. doi:10.1016/j.ajem.2010.11.029
104. Rabeneck L, Paszat LF, Li C. Risk Factors for Obstruction , Perforation , or Emergency Admission at Presentation in Patients with Colorectal Cancer : A Population-Based Study. 2006:1098-1103. doi:10.1111/j.1572-0241.2006.00488.x
105. Renzi C, Lyratzopoulos G, Card T, Chu TPC, Macleod U, Rachet B. Do colorectal cancer patients diagnosed as an emergency differ from non-emergency patients in their consultation patterns and symptoms? A longitudinal data-linkage study in England. *Br J Cancer*. 2016;115(August):1-10. doi:10.1038/bjc.2016.250
106. Gunnarsson H, Jennische K, Forssell S, et al. Heterogeneity of colon cancer patients reported as emergencies. *World J Surg*. 2014;38(7):1819-1826. doi:10.1007/s00268-014-2449-7
107. Office for National Statistics.. Cancer Registration Statistics. Cancer Registration Statistics.
108. Siegel RL, Fedewa SA, Anderson WF, et al. Colorectal Cancer Incidence Patterns in the

- United States, 1974–2013. *JNCI J Natl Cancer Inst.* 2017;109(8):27-32. doi:10.1093/jnci/djw322
109. O’Connell JB, Maggard MA, Liu JH, Etzioni DA, Livingston EH, Ko CY. Rates of colon and rectal cancers are increasing in young adults. *Am Surg.* 2003;69(10):866-872. <http://www.ncbi.nlm.nih.gov/pubmed/14570365>. Accessed July 25, 2017.
 110. You YN. Young-Onset Colorectal Cancer: Is It Time to Pay Attention? *Arch Intern Med.* 2012;172(3):287. doi:10.1001/archinternmed.2011.602
 111. Winchester DP, Stewart AK, Phillips JL, Ward EE. The National Cancer Data Base: Past, Present, and Future. *Ann Surg Oncol.* 2010;17(1):4-7. doi:10.1245/s10434-009-0771-3
 112. Inra JA, Syngal S. Colorectal Cancer in Young Adults. 2015:722-733. doi:10.1007/s10620-014-3464-0
 113. Wang R, Mo-jin W, Ping J. Clinicopathological Features and Survival Outcomes of Colorectal Cancer in Young Versus Elderly. 2015;94(35):1-8. doi:10.1097/MD.0000000000001402
 114. Kim T, Kim E, Hong S, Chang D, Kim Y-H. Long-Term Outcome and Prognostic Factors of Sporadic Colorectal Cancer in Young Patients. 2016;95(19):1-8. doi:10.1097/MD.0000000000003641
 115. Abdelsattar ZM, Wong SL, Regenbogen SE, Jomaa DM. Colorectal Cancer Outcomes and Treatment Patterns in Patients Too Young for Average-Risk Screening. 2016:929-934. doi:10.1002/cncr.29716
 116. Li M, Li J-Y, Zhao A-L, Gu J. Do Young Patients with Colorectal Cancer Have a Poorer Prognosis than Old Patients? *J Surg Res.* doi:10.1016/j.jss.2009.10.037
 117. Connell JBO, D M, Maggard MA, et al. Colorectal cancer in the young. 2004;187:343-348. doi:10.1016/j.amjsurg.2003.12.020
 118. 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(5):728-737.e3. doi:10.1016/j.cgh.2016.10.038
 119. Dozois EJ, Boardman LA, Suwanthanma W, et al. Young-onset colorectal cancer in patients with no known genetic predisposition: can we increase early recognition and improve outcome? *Medicine (Baltimore).* 2008;87(5):259-263. doi:10.1097/MD.0b013e3181881354
 120. Davis DM, Marcet JE, Frattini JC, Prather AD, Mateka JJL, Nfonsam VN. Is It Time to Lower the Recommended Screening Age for Colorectal Cancer? 2011;213(3):352-361. doi:10.1016/j.jamcollsurg.2011.04.033
 121. Herrett E, Gallagher AM, Bhaskaran K, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *Int J Epidemiol.* 2015;44(3):827-836. doi:10.1093/ije/dyv098
 122. MHRA. Clinical Practice Research Datalink. www.cprd.com/home/.
 123. Williams T, van Staa T, Puri S, Eaton S. Recent advances in the utility and use of the General Practice Research Database as an example of a UK Primary Care Data resource. *Ther Adv drug Saf.* 2012;3(2):89-99. doi:10.1177/2042098611435911
 124. 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(5):e011664. doi:10.1136/bmjopen-2016-011664
 125. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: A systematic review. *Br J Clin Pharmacol.* 2010;69:4-14. doi:10.1111/j.1365-2125.2009.03537.x
 126. Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: A systematic review. *Br J Gen Pract.* 2010;60(March):199-206. doi:10.3399/bjgp10X483562
 127. Dregan A, Moller H, Murray-Thomas T, Gulliford MC. Validity of cancer diagnosis in a primary care database compared with linked cancer registrations in England. Population-based cohort study. *Cancer Epidemiol.* 2012;36(5):425-429. doi:10.1016/j.canep.2012.05.013
 128. Boggon R, van Staa TP, Chapman M, Gallagher AM, Hammad TA, Richards MA. Cancer

- recording and mortality in the General Practice Research Database and linked cancer registries. *Pharmacoepidemiol Drug Saf.* 2013;22(2):168-175. doi:10.1002/pds.3374
129. Burns EM, Rigby E, Mamidanna R, et al. Systematic review of discharge coding accuracy. *J Public Health (Oxf)*. 2012;34(1):138-148. doi:10.1093/pubmed/fdr054
130. Public Health England. National Cancer Registration and Analysis Service. <http://www.ncin.org.uk/home>.
131. ONS. Cancer Registration Statistics for England. <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/methodologies/cancerregistrationstatisticsqmi#output-quality>. Accessed August 20, 2007.
132. ONS. User guide to mortality statistics.
133. Esteva M, Ruiz A, Ramos M, et al. Age differences in presentation, diagnosis pathway and management of colorectal cancer. *Cancer Epidemiol.* 2014;38:346-353. doi:10.1016/j.canep.2014.05.002
134. van der Geest LGM, Portielje JEA, Wouters MWJM, et al. Complicated postoperative recovery increases omission, delay and discontinuation of adjuvant chemotherapy in patients with Stage III colon cancer. *Colorectal Dis.* 2013;15(10):e582-91. doi:10.1111/codi.12288
135. Crooks CJ, West J, Card TR. *A Comparison of the Recording of Comorbidity in Primary and Secondary Care by Using the Charlson Index to Predict Short-Term and Long-Term Survival in a Routine Linked Data Cohort.* Vol 5.; 2015:e007974. doi:10.1136/bmjopen-2015-007974
136. Khan NF, Perera R, Harper S, Rose PW. Adaptation and validation of the Charlson Index for Read/OXMIS coded databases. *BMC Fam Pract.* 2010;11(1):1. doi:10.1186/1471-2296-11-1
137. Noble M, Wright G, Smith G, Dibben C. Measuring Multiple Deprivation at the Small-Area Level. *Environ Plan A.* 2006;38(1):169-185. doi:10.1068/a37168
138. Walters S, Maringe C, Butler J, Brierley JD, Rachet B, Coleman MP. Comparability of stage data in cancer registries in six countries: Lessons from the International Cancer Benchmarking Partnership. *Int J Cancer.* 2013;132(3):676-685. doi:10.1002/ijc.27651
139. NCIN. National Report on Colorectal Staging Data Quality Report, Colorectal Malignancies Site-Specific Clinical Reference Group.
140. Beral V, Peto R. UK cancer survival statistics. *BMJ.* 2010;341:c4112. doi:10.1136/BMJ.C4112
141. Herrett E, Thomas SL, Schoonen WM, Smeeth L, Hall AJ. Validation and validity of diagnoses in the General Practice Research Database: A systematic review. *Br J Clin Pharmacol.* 2010;69(1):4-14. doi:10.1111/j.1365-2125.2009.03537.x
142. Pollock AM, Vickers N. Trends in colorectal cancer care in southern England, 1989-1993: using HES data to inform cancer services reviews. *J Epidemiol Community Health.* 1998;52(7):433-438.
143. Gill S, Loprinzi CL, Sargent DJ, et al. Pooled analysis of fluorouracil-based adjuvant therapy for stage II and III colon cancer: Who benefits and by how much? *J Clin Oncol.* 2004;22(10):1797-1806. doi:10.1200/JCO.2004.09.059
144. Biagi JJ, Raphael MJ, Mackillop WJ, Kong W, King WD, Booth CM. Association between time to initiation of adjuvant chemotherapy and survival in colorectal cancer: a systematic review and meta-analysis. *JAMA.* 2011;305(22):2335-2342. doi:10.1001/jama.2011.749
145. Des Guetz G, Nicolas P, Perret G-Y, Morere J-F, Uzzan B. Does delaying adjuvant chemotherapy after curative surgery for colorectal cancer impair survival? A meta-analysis. *Eur J Cancer.* 2010;46(6):1049-1055. doi:10.1016/j.ejca.2010.01.020
146. Hershman D, Hall MJ, Wang X, et al. Timing of adjuvant chemotherapy initiation after surgery for stage III colon cancer. *Cancer.* 2006;107(11):2581-2588. doi:10.1002/cncr.22316
147. Tevis SE, Kohlnhofer BM, Stringfield S, et al. Postoperative complications in patients with rectal cancer are associated with delays in chemotherapy that lead to worse disease-free and overall survival. *Dis Colon Rectum.* 2013;56:1339-1348. doi:10.1097/DCR.0b013e3182a857eb

148. Nachiappan S, Askari A, Mamidanna R, et al. The impact of adjuvant chemotherapy timing on overall survival following colorectal cancer resection. *Eur J Surg Oncol*. 2015;41(12):1636-1644. doi:10.1016/j.ejso.2015.09.009
149. Bouras G, Burns EM, Howell A-M, Bottle A, Athanasiou T, Darzi A. Risk of Post-Discharge Venous Thromboembolism and Associated Mortality in General Surgery: A Population-Based Cohort Study Using Linked Hospital and Primary Care Data in England. *PLoS One*. 2015;10(12):e0145759. doi:10.1371/journal.pone.0145759
150. Ahmed S, Ahmad I, Zhu T, et al. Early discontinuation but not the timing of adjuvant therapy affects survival of patients with high-risk colorectal cancer: a population-based study. *Dis Colon Rectum*. 2010;53(10):1432-1438. doi:10.1007/DCR.0b013e3181e78815
151. Bayraktar UD, Chen E, Bayraktar S, et al. Does delay of adjuvant chemotherapy impact survival in patients with resected stage II and III colon adenocarcinoma? *Cancer*. 2011;117(11):2364-2370. doi:10.1002/cncr.25720
152. Cheung WY, Neville BA, Earle CC. Etiology of delays in the initiation of adjuvant chemotherapy and their impact on outcomes for stage II and III rectal cancer. *Dis Colon Rectum*. 2009;52(6):1054-1063. doi:10.1007/DCR.0b013e3181a51173
153. Wadd N, Borowski D, Bharathan B, Kelly S, Bradburn D WR. Timing of adjuvant chemotherapy in Dukes' stage C colorectal cancer: does a delayed start of treatment affect outcome? *Ann Oncol*. 2008;19(suppl 6):24.
154. Partnership HQL. National Bowel Cancer Audit. <http://content.digital.nhs.uk/bowel>. Accessed November 9, 2016.
155. Cheung WY, Neville BA, Earle CC. Etiology of delays in the initiation of adjuvant chemotherapy and their impact on outcomes for Stage II and III rectal cancer. *Dis Colon Rectum*. 2009;52(6):1054-1063; discussion 1064. doi:10.1007/DCR.0b013e3181a51173
156. Raymond TM, Dastur JK, Khot UP, Parker MC. Hospital Stay and Return to Full Activity Following Laparoscopic Colorectal Surgery. *JSLS J Soc Laparoendosc Surg*. 2008;12(2):143-149.
157. Nachiappan S, Askari A, Mamidanna R, et al. Initiation of adjuvant chemotherapy within 8 weeks of elective colorectal resection improves overall survival regardless of reoperation. *Color Dis*. 2016;18(11):1041-1049. doi:10.1111/codi.13308
158. Sharma A, Deeb A-P, Iannuzzi JC, Rickles AS, Monson JRT, Fleming FJ. Tobacco Smoking and Postoperative Outcomes After Colorectal Surgery. *Ann Surg*. 2012;258(2):296-300. doi:10.1097/SLA.0b013e3182708cc5
159. Sanoff HK, Carpenter WR, Martin CF, et al. Comparative effectiveness of oxaliplatin vs non-oxaliplatin-containing adjuvant chemotherapy for stage III colon cancer. *J Natl Cancer Inst*. 2012;104(3):211-227. doi:10.1093/jnci/djr524
160. Morris M, Platell C, Fritschi L, Iacopetta B. Failure to complete adjuvant chemotherapy is associated with adverse survival in stage III colon cancer patients. *Br J Cancer*. 2007;96(5):701-707. doi:10.1038/sj.bjc.6603627
161. Lyratzopoulos G, Saunders CL, Abel GA, et al. The relative length of the patient and the primary care interval in patients with 28 common and rarer cancers. *Br J Cancer*. 2015;112 Suppl(s1):S35-40. doi:10.1038/bjc.2015.40
162. Neal RD, Allgar VL, Ali N, et al. Stage, survival and delays in lung, colorectal, prostate and ovarian cancer: Comparison between diagnostic routes. *Br J Gen Pract*. 2007;57(536):212-219.
163. 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(10):977-981. doi:10.1016/j.ejso.2010.07.002
164. Keeble S, Abel GA, Saunders CL, et al. Variation in promptness of presentation among 10,297 patients subsequently diagnosed with one of 18 cancers: evidence from a National Audit of Cancer Diagnosis in Primary Care. *Int J cancer*. 2014;135(5):1220-1228.

- doi:10.1002/ijc.28763
165. 2014 National Cancer Patient Experience Survey. <https://www.quality-health.co.uk/resources/surveys/national-cancer-experience-survey/>.
 166. Neal RD, Tharmanathan P, France B, et al. Is increased time to diagnosis and treatment in symptomatic cancer associated with poorer outcomes? Systematic review. *Br J Cancer*. 2015;112(March):S92-S107. doi:10.1038/bjc.2015.48
 167. Iversen LH, Antonsen S, Laurberg S, Lautrup MD. Therapeutic delay reduces survival of rectal cancer but not of colonic cancer. *Br J Surg*. 2009;96:1183-1189. doi:10.1002/bjs.6700
 168. Stapley S, Peters TJ, Sharp D, Hamilton W. The mortality of colorectal cancer in relation to the initial symptom at presentation to primary care and to the duration of symptoms: a cohort study using medical records. *Br J Cancer*. 2006;95(10):1321-1325. doi:10.1038/sj.bjc.6603439
 169. Van Hout AMGH, de Wit NJ, Rutten FH, Peeters PHM. Determinants of patient's and doctor's delay in diagnosis and treatment of colorectal cancer. *Eur J Gastroenterol Hepatol*. 2011;1. doi:10.1097/MEG.0b013e32834c4839
 170. Alepuz MT, Gonzalez D, Martín M. Impact on patient care time and tumor stage of a program for fast diagnosis and treatment of colorectal cancer. 2011;103:13-19.
 171. Murchie P, Raja E a, Brewster DH, 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(3):1-9. doi:10.1038/bjc.2014.352
 172. 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(7602):1040. doi:10.1136/bmj.39171.637106.AE
 173. Alexiusdottir KK, Möller PH, Snaebjornsson P, et al. Association of symptoms of colon cancer patients with tumor location and TNM tumor stage. *Scand J Gastroenterol*. 2012;47(7):795-801. doi:10.3109/00365521.2012.672589
 174. Korsgaard M, Pedersen L, Sørensen HT, Laurberg S. Reported symptoms, diagnostic delay and stage of colorectal cancer: A population-based study in Denmark. *Color Dis*. 2006;8:688-695. doi:10.1111/j.1463-1318.2006.01014.x
 175. Dregan A, Möller H, Charlton J, Gulliford MC. Are alarm symptoms predictive of cancer survival?: population-based cohort study. *Br J Gen Pract*. 2013;63(617):e807-12. doi:10.3399/bjgp13X675197
 176. Ben-Ishay O. Clinical presentation predicts the outcome of patients with colon cancer. *World J Gastrointest Surg*. 2013;5(4):104. doi:10.4240/wjgs.v5.i4.104
 177. Thompson MR, Perera R, Senapati A, Dodds S. Predictive value of common symptom combinations in diagnosing colorectal cancer. *Br J Surg*. 2007;94(10):1260-1265. doi:10.1002/bjs.5826
 178. UK CR. National Cancer Diagnosis Audit | Cancer Research UK. http://www.cancerresearchuk.org/health-professional/early-diagnosis-activities/national-cancer-diagnosis-audit?utm_campaign=NAEDI_maynewsletter_310517_ccmpe&utm_source=ccmp&utm_medium=email&cid=59128&mid=147424490. Accessed June 29, 2017.
 179. Ahnen DJ, Wade SW, Jones WF, et al. The Increasing Incidence of Young-Onset Colorectal Cancer: A Call to Action. *Mayo Clin Proc*. 2014;89(2):216-224. doi:10.1016/j.mayocp.2013.09.006
 180. Law J, Koh FH, Tan K. Young colorectal cancer patients often present too late. 2017;8-12. doi:10.1007/s00384-017-2837-1
 181. Downing A, Aravani A, Macleod U, et al. Early mortality from colorectal cancer in England : a retrospective observational study of the factors associated with death in the first year after diagnosis. *Br J Cancer*. 2013;108(3):681-685. doi:10.1038/bjc.2012.585
 182. Hamilton W. Emergency admissions of cancer as a marker of diagnostic delay. *Br J Cancer*. 2012;107(8):1205-1206. doi:10.1038/bjc.2012.423
 183. Bailey CE, Hu C-Y, You YN, et al. Increasing Disparities in the Age-Related Incidences of

- Colon and Rectal Cancers in the United States, 1975-2010. *JAMA Surg.* 2015;150(1):17. doi:10.1001/jamasurg.2014.1756
184. Lyratzopoulos G, Neal RD, Barbiere JM, Rubin GP, Abel GA. Variation in number of general practitioner consultations before hospital referral for cancer: Findings from the 2010 National Cancer Patient Experience Survey in England. *Lancet Oncol.* 2012;13(4):353-365. doi:10.1016/S1470-2045(12)70041-4
 185. Mitchell ED, Pickwell-Smith B, Macleod U. Risk factors for emergency presentation with lung and colorectal cancers: a systematic review. *BMJ Open.* 2015;5(4):e006965. doi:10.1136/bmjopen-2014-006965
 186. Hamilton W, Green T, Martins T, Elliott K, Rubin G, Macleod U. Evaluation of risk assessment tools for suspected cancer in general practice: a cohort study. *Br J Gen Pract.* 2013;63(606):e30-6. doi:10.3399/bjgp13X660751
 187. Beggs AD, Bhate RD, Irukulla S, Achiek M, Abulafi AM. Straight to colonoscopy : the ideal patient pathway for the 2-week suspected cancer referrals ? 2011:114-119. doi:10.1308/003588411X12851639107917
 188. Libby G, Brewster DH, McClements PL, et al. The impact of population-based faecal occult blood test screening on colorectal cancer mortality: a matched cohort study. *Br J Cancer.* 2012;107(2):255-259. doi:10.1038/bjc.2012.277
 189. Liu Z, Zhang Y, Niu Y, et al. A systematic review and meta-analysis of diagnostic and prognostic serum biomarkers of colorectal cancer. *PLoS One.* 2014;9(8):e103910. doi:10.1371/journal.pone.0103910
 190. Office for National Statistics. *Deaths Registered in England and Wales 2014.*; 2014. http://www.ons.gov.uk/ons/dcp171778_410352.pdf.
 191. UK CR. Cancer Mortality Statistics. <http://www.cancerresearchuk.org/health-professional/cancer-statistics/mortality#heading-Five>.
 192. Berrino F, De Angelis R, Sant M, et al. Survival for eight major cancers and all cancers combined for European adults diagnosed in 1995-99: results of the EURO CARE-4 study. *Lancet Oncol.* 2007;8(9):773-783. doi:10.1016/S1470-2045(07)70245-0
 193. Li R, Abela L, Moore J, et al. Control of data quality for population-based cancer survival analysis. *Cancer Epidemiol.* 2014;38(3):314-320. doi:10.1016/j.canep.2014.02.013
 194. Gatta G, Faivre J, Capocaccia R, Ponz de Leon M. Survival of colorectal cancer patients in Europe during the period 1978-1989. EURO CARE Working Group. *Eur J Cancer.* 1998;34(14 Spec No):2176-2183. <http://www.ncbi.nlm.nih.gov/pubmed/10070284>. Accessed June 1, 2015.
 195. Pollock AM, Vickers N. Why are a quarter of all cancer deaths in south-east England registered by death certificate only? Factors related to death certificate only registrations in the Thames Cancer Registry between 1987 and 1989. *Br J Cancer.* 1995;71(3):637-641.
 196. Dickinson HO, Salotti JA, Birch PJ, Reid MM, Malcolm A, Parker L. How complete and accurate are cancer registrations notified by the National Health Service Central Register for England and Wales? *J Epidemiol Community Health.* 2001;55(6):414-422.
 197. Hiom SC. Diagnosing cancer earlier: reviewing the evidence for improving cancer survival. *Br J Cancer.* 2015;(March):1-5. doi:10.1038/bjc.2015.23
 198. Amri R, Bordeianou LG, Sylla P, Berger DL. Treatment Delay in Surgically-Treated Colon Cancer: Does It Affect Outcomes? *Ann Surg Oncol.* 2014;21(12):3909-3916. doi:10.1245/s10434-014-3800-9
 199. Gatta G, Zigon G, Aareleid T, et al. Patterns of care for European colorectal cancer patients diagnosed 1996-1998: a EURO CARE high resolution study. *Acta Oncol.* 2010;49(6):776-783. doi:10.3109/02841861003782009
 200. Pescatori M, Maria G, Beltrani B, Mattana C. Site, emergency, and duration of symptoms in the prognosis of colorectal cancer. *Dis Colon Rectum.* 1982;25(1):33-40.
 201. Walter F, Webster A, Scott S, Emery J. The Andersen Model of Total Patient Delay: a systematic review of its application in cancer diagnosis. *J Health Serv Res Policy.* 2012;17(2):110-118. doi:10.1258/jhsrp.2011.010113

202. Esteva M, Leiva A, Ramos M, et al. Factors related with symptom duration until diagnosis and treatment of symptomatic colorectal cancer. *BMC Cancer*. 2013;13:87. doi:10.1186/1471-2407-13-87
203. Taylor EF, Thomas JD, Whitehouse LE, et al. Population-based study of laparoscopic colorectal cancer surgery 2006-2008. *Br J Surg*. 2013;100(4):553-560. doi:10.1002/bjs.9023
204. Dregan A, Møller H, Charlton J, Gulliford MC. Are alarm symptoms predictive of cancer survival? *Br J Gen Pract*. 2013;63(617):807-812. doi:10.3399/bjgp13X675197
205. Barrett J. Pathways to the diagnosis of colorectal cancer: an observational study in three UK cities. *Fam Pract*. 2005;23(1):15-19. doi:10.1093/fampra/cmi093
206. Khattak I, Eardley NJ, Rooney PS. Colorectal cancer - A prospective evaluation of symptom duration and GP referral patterns in an inner city teaching hospital. *Color Dis*. 2006;8(6):518-521. doi:10.1111/j.1463-1318.2006.00967.x
207. Burns EM, Bottle A, Aylin P, Darzi A, Nicholls RJ, Faiz O. Variation in reoperation after colorectal surgery in England as an indicator of surgical performance: retrospective analysis of Hospital Episode Statistics. *BMJ*. 2011;343:d4836. doi:10.1136/bmj.d4836

Appendix

i Study protocol submitted to CPRD and ISAC to obtain data for thesis

A. Lay Summary

Cancer is a major cause of mortality in the UK, accounting for nearly one in three deaths. Survival following diagnosis in the UK is lower compared to other Western and Northern European countries, with the largest difference in survival occurring in the first year after diagnosis. Also a higher percentage of patients present as an emergency in the UK. We believe delays along the patient's pathway accounts for these differences. CPRD linked data offers a unique opportunity to define all the clinical episodes along the pathway from first GP presentation to long term follow up including surgical and chemo- or radiotherapy treatment episodes. Our population will include all patients diagnosed with colorectal, breast, prostate, oesophageal, gastric or lung cancer. The findings of this study can help formulate a public health policy that concentrates resources along the parts of the pathway that are most likely to improve survival in the UK.

B. Technical Summary

Survival following cancer diagnosis in the UK is lower compared to other Western and Northern European countries. The symptomatic patient pathway begins with the first symptom, and involves clinical events such as the first GP consultation, hospital appointments, investigations, and treatment. Our hypothesis is that detrimental events during the patients diagnosis and treatment lead to delays, causing the poor 1-year survival seen in the UK. Our aim is to identify the causes of these delays by analysing CPRD linked data - GP consultations, HES admissions, NCIN cancer dataset and ONS mortality data. CPRD is the largest source of routinely collected information available in the UK. Studies have shown CPRD is a reliable source of consultation dates, co-morbidities, complications, medications and cancer diagnoses. Further information regarding chemo-/radiotherapy and complications following surgery can be identified in HES. However, CPRD linked data is limited as the length of symptoms before consultation and outpatient data is not routinely available.

For the purpose of this study, the patient pathway will be split into before and after diagnosis. We will determine the effect of the type of symptom, co-morbidity, demographics and geographical location on a delay in diagnosis, and how this effects survival. Following diagnosis, the type of operation (if relevant), stage of tumour, co-morbidities and complications will be used in multilevel regression modelling to determine the effect on survival.

The date of diagnosis will be determined on a hierarchal level of evidence as set out by the International Association of Cancer Registries, with first histological confirmation being the preferred date of diagnosis.

Our study population criterion is any patient over the age of 18 that has a diagnosis of colorectal, breast, prostate, oesophogastric or lung cancer in the NCIN cancer dataset, and has linked data in CPRD.

Patients who survive beyond one year after diagnosis will act as controls, as previous studies have shown the largest difference in survival between the UK and northern Europe occurs in the first 12 months. Secondary analysis will compare patients who have had a curative resection to those who have not, as the latter will be used as a surrogate for palliative treatment. A third analysis will compare patients diagnosed electively vs. emergency, as the latter group are known to have a worse survival. Previous studies that have attempted to test this hypothesis using local datasets or notes have not been able to include sufficient patient numbers and confounders.

The accuracy of HES has been previously verified, by comparing to patient notes. Analysis of the CPRD data from within our department has demonstrated there is sufficient number of clinical episodes

regarding cancer symptoms before diagnosis, and complications in the community following treatment. A similar process has also been carried out for CPRD and is described in a systematic review. This latter publication found a median of over 95% (range 74 – 100) concordance for cancer related diagnoses. A concordance of 83% has also been noted when CPRD has been compared to the cancer registry. Although there remains a possibility of missing clinical episodes, these publications suggest such occurrences should be low.

C. Objectives, Specific Aims and Rationale

(i) Objective

To use the GP database, NCIN cancer registry, HES and ONS cause of death data to determine the delays along the cancer patient pathway that lead to a poor outcome in the UK. Our study population includes all patients with an ICD diagnosis of colorectal, breast, prostate, oesophogastric or lung cancer in the NCIN dataset.

(ii) Hypothesis and aims

Our hypothesis: delays along the cancer patient pathway account for the poor outcome noted in the UK. We are aiming to

- Compare the pathways between patients who die within one year of diagnosis to those that survive beyond one year to determine the cause and effect of delays
- Establish the factors along the pathway that leads a patient to be treated with palliative rather than with curative intent
- Determine the reasons why patients present as an emergency rather than electively
- Describe the effect of the type of operation and complications on 1-year and conditional 5 year survival
- Use regression analysis to determine the confounders and causes of delays

D. Background

Cancer is a major cause of mortality in the UK, accounting for nearly one in three deaths¹⁹⁰. 46% of all deaths from cancer are attributed to lung, breast, bowel and prostate¹⁹¹. Despite a number of national initiatives, survival in the UK following a cancer diagnosis is lower compared to other Western and Northern European countries¹³⁸. The disparity was first noted by the Eurocare-1 study⁶, which compared the national cancer registries for ten European Countries. This study concentrated on new cases of colorectal cancers between 1978 and 1985. It demonstrated the relative risk of mortality in the first 6 months of diagnosis was higher in the UK compared to Finland (1.58 (CI 1.39 to 1.79))⁶. Survival beyond the first 6 months was also worse. Analysis of subsequent Eurocare data by Abdel-Rahman et al.⁷ suggested 6600 – 7500 excess cancer related deaths could have been avoided in the UK every year leading up to 1999 if the UK had matched the European average. This increases to 11400 compared to the best performing European countries. More recent analysis of cancer registry data by the EURO CARE group^{11,192} and the International Cancer Benchmarking Partnership (ICBP)¹², has once again demonstrated that the 1-year and 5-year survival is worse in the UK.

Various theories have been suggested to account for these differences, including disparity in the method and accuracy of cancer registry data collection^{9,10,21,193}. The UK was represented by only four regional cancer registries in the initial EURO CARE studies, one of which had a relatively high percentage of missing data^{194–196}. However the latter studies by EURO CARE and ICBP^{11,12}, as mentioned above, have utilised data from all the UK cancer registries, all of which conform to an international standard.

The largest difference in survival from cancer occurs within the first year after diagnosis^{12,22,197}. The conditional 5-year survival in the UK for all cancers is closer to the better performing European countries

than the 1-year survival^{12,22}. Although surgery is the preferred treatment choice for colorectal, breast and adenocarcinoma of the oesophagus and stomach, a higher percentage of patients within the UK are treated without surgery compared to the better performing European countries¹⁸. In addition, there are a relatively higher number of emergency cancer diagnoses in the UK¹⁰⁰. For the equivalent stage, emergency patients have a worse outcome compared to elective patients^{29,90,100}. It is unclear why UK patients are not identified earlier in their pathway to allow elective treatment with curative intention.¹⁹⁷.

These findings indicate the significant events that determine long-term survival should be found in the 1st year following start of symptoms. Possible factors include delays in treatment^{198,199}, delays in diagnosis^{30,79,81,197}, co-morbidities²⁷, complications following resection^{134,147}, mode of presentation (emergency vs elective)^{90,200} or tumour stage at diagnosis^{19,21}. Studies to date have failed to include sufficient patients and confounders for multivariate regression analysis to conclusively demonstrate which factors are significant.

The patient pathway is complex and can be split into various stages for the purpose of analysis. The pathway for symptomatic patients begins with the first sign or symptom that leads the patient to seek medical attention²⁰¹, and ends with 5-year disease free survival or death³⁰. Majority of this pathway occurs in the community, where the primary care physician acts as the gatekeeper for care.

The word 'delay' is often used by authors as a narrative description of the time elapsed between two events in the patient pathway, even though the time elapsed may not be detrimental. For example zero day 'delay' between the first presentation to GP and the referral to the hospital. There is no consensus as to how many days constitutes a detrimental delay between the time points mentioned below. A better term may be "time elapsed" as has previously been suggested²⁰².

The HES and CPRD databases have been used separately to map out the initial pathway to diagnosis^{814,102,166,203,204}. CPRD linkage has been used for diagnosis, but this analysis did not include palliative or emergency patients⁸¹. Ellis et al. used HES to identify the method of admission²⁹. However CPRD GOLD, HES, the NCIN cancer registry and cause of death from the Office for National Statistics has not been previously used together to identify each important event along the patient pathway from first symptom to death or long term follow up.

Other methods have been used to identify the cause and effect of delays in the pathway for cancers. Patient and GP surveys are able to identify the start type and timing of symptoms^{64,174}. However this method introduces recall bias, which varies according to how far along the pathway is the patient at the time of the survey. Alternatively authors have carried out retrospective analysis of hospital and primary care notes^{41,205}. This method is time consuming and expensive, limiting the number of cases that can be included. This results in insufficient numbers to produce statistically significant conclusions. Local databases, with contemporaneously collected information, have also been used, but these represent local practices that may not be relevant to the rest of the UK²⁰⁶. Retrospective notes analysis has been used to determine the optimum timing of chemo- or radiotherapy¹⁹⁸. Alternatively clinical practice regarding the timing of neo-adjuvant or adjuvant treatment may be based on sub-group analysis of data from trials which were not designed for this purpose¹⁴⁴.

E. Study Type

The study will primarily be hypothesis testing.

F. Study Design

This is a population-based retrospective cohort study.

Studies have shown 1-year survival in the UK is worse than most other Western European countries. Therefore, a cohort study will allow comparison of pathways between patients who die within 1 year compared to those who survive longer. In addition, if there are sufficient numbers, we would also analyse the 5-year conditional survival.

G. Sample Size

Based on previous studies, the following number of patients can be identified using the NCIN dataset, if inclusion is restricted to those with at least one related symptom Medcode recorded in CPRD within one year of diagnosis: 8500 breast patients, 6000 colorectal patients, 5500 lung, 1700 prostate, 2000 gastro-oesophageal⁸¹¹²⁷. These figures are for diagnoses up to 2009. We would expect higher numbers as the NCIN dataset has been extended to end of 2013, and we have requested all patients with a cancer diagnosis in NCIN dataset irrespective of presentations to GP. Analysis within our department has shown 14500 colorectal cancer patients can be identified from HES for patients with a linkage to CPRD.

Studies have shown a good concordance of cancer diagnoses between CPRD and cancer registry. Therefore, patients without a previous symptom Medcode in CPRD will be analysed separately. They may represent a cohort of emergency treatment patients, but this is to be determined. Also by obtaining all consecutive cases covered in the NCIN dataset rather than a random sample of cancer patients will allow for meaningful multi-level hierarchical modelling.

Mr G. Bouras (PhD student within our department) found that there was sufficient number of patients in CPRD to carry out a multi-level regression analysis for 981 patients with a diagnosis of PE or DVT following surgery, whilst showing a statistical significant difference ($p < 0.05$). Factors included were type of surgery (elective vs emergency), age, BMI, length of hospital stay, stage of cancer, re-operations and co-morbidities.¹⁴⁹

For our latest study, we will be analyzing these same factors and in addition chemo-/radiotherapy episodes in HES. Therefore, we believe CPRD will provide enough numbers and detail to fulfil our aims.

H. Data Linkage Required (if applicable)

To test our hypothesis, it is necessary to include all clinical episodes encountered by the patient in primary, CPRD, and secondary care, HES. Previous studies have utilised one or the other, therefore not providing enough confounders to determine the cause and effect of delays.

The following factors will be extracted for analysis and tested as covariates:

- Time elapsed between clinically important events as described below (HES and CPRD)
- Type of symptoms and frequency of GP consultations before diagnosis (CPRD)¹²³
- Frequency of GP presentations for non-cancer related consultations
- Type of presentation – elective vs emergency (HES)²⁹
- Type of operation (HES), and complications (HES and CPRD)²⁰⁷
- Palliative or Curative intent (HES and CPRD)
- Patient demographics such as age, co-morbidities, BMI, geographical location, smoking status (CPRD)
- Socioeconomic status (Townsend score)
- Tumour stage (NCIN)⁹
- Chemo- and/or Radiotherapy treatment (HES) – both adjuvant and neo-adjuvant¹⁴⁸
- Cause of death – cancer and non- cancer specific (ONS)¹⁰².

We understand the latest available coverage periods for datasets are:

Cancer Registry 01/0/1990 to 31/12/2013
Inpatient HES data -1/04/1997 to 31/03/2015
ONS mortality data – 01/01/1998 to 31/03/2015

I. Study Population

The population is all consecutive patients with an ICD-10 code of colorectal, breast, lung, prostate, oesophagus or stomach cancer (Appendix A) in the NCIN cancer registry. This dataset includes all patients that have been registered with a diagnosis of cancer from 1997 to December 2013. The dataset will be restricted by the coverage of linked GP data from 1997 to latest available update. We will require at least 1 year follow-up.

Experience within our research group of using the cancer registry has shown the data has sufficient completeness to carry out this project. For example, stage data for colorectal cancer is about 80% complete, which increase with the more current registry data. Information from the CPRD website suggests 83.3% concordance of diagnosis in GOLD and NCIN. Therefore, we have used the latter as the source.

By utilising the cancer registry as the source, we will identify patients who were treated without a curative resection. For example, patients with advanced colorectal cancer, or hormone treatment for elderly patients and/or advanced disease. For these particular patients, we are aiming to identify the number of GP consultations, the type of symptom and/or co-morbidities that may have delayed the diagnosis or treatment.

J. Selection of comparison group(s) or controls

Comparisons will be made between patients who die within 1 year and a control group of those that survive beyond one year. This is because the outcome for UK cancer patients is particularly worse in the first year.

Secondary analysis would be to compare patients based on intention to treat:- resection and palliative. Surgery is the preferred treatment option for solid tumours, as it confers the best survival benefit. For the purpose of this study, a patient would be considered palliative if they 1) do not undergo a curative resection 2) are diagnosed stage IV disease 3) have an ICD code in HES of metastases without subsequent resection of metastases 4) have a Medcode in the GP database indicating palliative treatment 5) have a OPCS code for an operation indicating palliative treatment only e.g. oesophageal stent.

Thirdly, comparison will be between elective and emergency presentations. Elective treatment is preferred for all cancers, as it has the best outcome in terms of survival and reduced complications.

K. Exposures, Outcomes and Covariates

The following READ codes can help identify cancer related consultations within the GP database: oesophageal B10., stomach B11., colon B13., rectum B14., lung B22., breast B34.. &B35., prostate B46... The consultations before diagnosis are not expected to include the relevant READ cancer code. The GP database will provide events in the community such as consultations before referral to hospital and complications managed by the primary care physician. The relevant READ codes will be obtained from the NHS Information Authority Clinical Terminology Browser. The coding and dates for cancer diagnosis in primary and secondary care data will be compared.

According to the literature, delays between the following events in the cancer patient pathway may determine survival:

- Complications following surgery¹⁴⁷
- Start of symptoms related to cancer^{27,32,202}
- First presentation to the GP or Hospital due to these symptoms^{29,52}
- Date of diagnosis¹⁸
- First treatment whether neo-adjuvant chemo/radio- therapy or surgery^{19,167}
- Subsequent treatment in the form of surgery or adjuvant chemoradiotherapy.^{144,146}

Complications: Complications are key clinical events that can cause delays to further treatment following surgery. Certain complications, such as an anastomotic leak, can reduce disease free survival. As noted by a George Bouras, a PhD student in our department, the GP database is an important source of complications that are not noted in the HES database (ISAC13_078R). The effect of these community managed complications has not been previously considered in terms of delays in adjuvant treatment and long term outcome.

Start of symptoms: It is not possible to accurately determine the start of symptoms using CPRD, as there are no READ or ICD codes that describe the duration of symptoms in sufficient detail. Therefore, for the purpose of this study, the patient pathway begins with the first presentation to the GP or Hospital with cancer related symptoms.

First presentation to GP: Appendix B contains the Medcodes codes for symptoms related to colorectal cancer, which will be used to identify the relevant GP consultations. We will determine the date of the first and all subsequent consultations before the patient was diagnosed. A similar table of Medcodes will be constructed for the other cancers.

First presentation to first hospital attendance: The ICD-10 codes for each cancer in Appendix A will be used to identify the relevant HES episodes. The method of admission is recorded in HES that determines whether the episode was an emergency presentation or an elective admission for treatment. Ellis et al. have shown that it is possible to use routinely collected national data to determine the method of presentation.

Diagnosis date: The Aarhus statement is a consensus agreement between experts in cancer pathway analysis that sets out the definition of important events used in this study¹. This states that the date of diagnosis should be equivalent to the date when the first histological sample is taken that reveals the cancer diagnosis. The hierarchy of evidence for the date of diagnosis is based on the European Association of Cancer Registry guidelines, and is available within the NCDR dataset^{9,10}. The Office of Population Census and Survey (OPCS) coding (version 4.2-4.6) will be used to identify the operation or investigation within HES that corresponds to the date of diagnosis in the NCDR dataset. This episode will not necessarily have an ICD code indicating cancer, as the diagnosis may not have been known at the time of the procedure.

First treatment: Following diagnosis, a decision by the consultant and MDT is made as to whether the patient should be treated with curative or palliative intent.

Subsequent treatment: This may be in the form of surgery or adjuvant chemo- or radiotherapy. Oral hormonal treatment for prostate and breast cancer can be determined from the prescription READ codes in the GP database.

End of patient pathway: The patient pathway ends at the date of death as recorded in the ONS dataset, or the end of data coverage, whichever occurs first. 1 year survival will be determined. The cause of death will be differentiated between cancer and non-cancer related.

L. Data/ Statistical analysis

All analyses will be performed on SPSS (version 22.0.) and Stata (Stata Corporation). Descriptive analyses of the study population will include the number of GP practices, years that the practices have participated in ~~CPRD~~ CPRD, deprivation, and geographical location of the practices. The precision of incidence calculations will be indicated by the reporting of 95% confidence intervals. Delays will be correlated against risk factors (age, sex, socioeconomic status, BMI, cancer stage and Charlson score), mortality and process measures (number of GP and hospital consultations). The type of risk factor tested will change according to the part of the pathway being analysed. For example, following rectal resection, the timing and number of cycles of adjuvant chemoradiotherapy, together with the influence of complications and co-morbidities on outcome. Univariate analysis will be performed using t-tests, chi-squared tests, Wilcoxon sum-rank tests as appropriate. Odds ratios and c-statistics will be reported where appropriate.

Publications to date have suggested survival following cancer diagnosis is multifactorial. Therefore, multi-level regression analysis will be utilised to find the best fit model to explain the findings. Survival, at one or five years, will be considered the outcome variable. The risk factors will be adjusted and tested according to the part of the pathway being analysed. For example, the number of GP consultations, socioeconomic status, type of symptom and type of presentation will be considered as explanatory variables before diagnosis. The same variables, in addition to type of operation, stage and complications will be used for the pathway following diagnosis. Additional factors are listed in section H.

Clinical events representing diagnoses and operations can occur at any point during the 15-year coverage of CPRD. Therefore, follow up is not uniform. An alternative method to analyse survival is to use Kaplan Meier survival analysis. This method censors patients who reach the end of ONS coverage without a death event. An example of such a technique would be to split patients into groups based on the time elapsed between surgery and first adjuvant chemotherapy treatment. The survival curves of these different cohorts will then be compared.

Plan for addressing confounding

The reliability and robustness of created regression models will be examined through sensitivity analyses. Multilevel hierarchical modelling will cluster patients, practices and secondary care providers. In addition, the following factors will be included in multilevel modelling: age, sex, co-morbidities, smoking history, tumour stage, tumour site, number of consultations, type of operation, use of stent (with colorectal) and neo-/adjuvant therapy. Survival analysis will be performed through multilevel Cox regression modelling and/or Kaplan Meier curves. Censoring will take into consideration patients who are considered alive at the end of data coverage.

M. Plan for addressing missing data

Using linked databases will potentially reduce the level of missing data as there is more than one source of information. Part of our project will be to compare the the dates and codes of diagnosis/operations between the datasets. This will reduce the potential bias of restricting analysis to one source.

Imputation methods will be used to complete data for tumour stage, hormone therapy and neo-/adjuvant treatment. Nevertheless, stage data is available for colorectal cancer in 80% of cases who have had a resection. This will provide about 11000 patients for analysis. We would expect further missing data for certain patient characteristics. For example, BMI is recorded in 60% of patients. Prescription information

is well documented in CPRD and will be used as a marker for a disease, such as Metformin for diabetes. Sensitivity analyses will be used to test the effect of missing data on regression models.

Patients who have left the practice will be identified using the 'transfer out date' code (tod) in CPRD. Out of these patients, those who have died will be identified using the transfer out reason code (toreason) and included in the analysis. The rest of the patients will be excluded if they have transferred from their practice within 6 months of their diagnosis as noted in NCIN.

N. Limitations of the study design, data sources and analytical methods

- This study relies on the detail recorded by the GP at the time of consultation
- The CPRD contains data derived from an existing coding system that was not designed specifically to measure patient delays.
- The timing of the start of symptoms is not reliably available through GP READ codes
- The effect of missing data may produce bias or equivocal results. Therefore, management of missing data needs to be carefully considered.

O. Patient or user group involvement (if applicable) – Not applicable

P. Plans for disseminating and communicating study results, including the presence or absence of any restrictions on the extent and timing of publication

The results of this research will be disseminated through peer-review publications, conference abstracts, presentations and media release.

We will take steps to prevent inadvertent disclosure, including restricting analysis to groups with at least 5 patients.

ii. Search strategy for review of CPRD accuracy

- 1 (general practice research database or GPRD).mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, nm, kf, px, rx, an, ui] (2696)
- 2 (Community Practice research datalink or CPRD).mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, nm, kf, px, rx, an, ui] (617)
- 3 1 or 2 (3291)
- 4 valid*.mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, nm, kf, px, rx, an, ui] (1197477)
- 5 consisten*.mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, nm, kf, px, rx, an, ui] (1499061)
- 6 sensitiv*.mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, nm, kf, px, rx, an, ui] (2954890)
- 7 specificity.mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, nm, kf, px, rx, an, ui] (1572954)
- 8 reliability.mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, nm, kf, px, rx, an, ui] (314289)
- 9 concordance.mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, nm, kf, px, rx, an, ui] (76911)
- 10 accuracy.mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, nm, kf, px, rx, an, ui] (853173)
- 11 variation.mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, nm, kf, px, rx, an, ui] (958497)
- 12 variab*.mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, nm, kf, px, rx, an, ui] (1964557)
- 13 verify*.mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, nm, kf, px, rx, an, ui] (122533)
- 14 recording*.mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, nm, kf, px, rx, an, ui] (420107)
- 15 (positive predictive value* or PPV*).mp. [mp=ti, ab, hw, tn, ot, dm, mf, dv, kw, nm, kf, px, rx, an, ui] (95485)
- 16 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 (9054609)
- 17 3 and 16 (1010)

iii ICD 10 codes for cancer diagnosis

ICD code	Colorectal
C18.0	Malignant neoplasm of cecum
C18.1	Malignant neoplasm of appendix
C18.2	Malignant neoplasm of ascending colon
C18.3	Malignant neoplasm of hepatic flexure
C18.4	Malignant neoplasm of transverse colon
C18.5	Malignant neoplasm of splenic flexure
C18.6	Malignant neoplasm of descending colon
C18.7	Malignant neoplasm of sigmoid colon
C18.8	Malignant neoplasm of overlapping sites of colon
C18.9	Malignant neoplasm of colon, unspecified
C19	Malignant neoplasm of rectosigmoid junction
C20	Malignant neoplasm of rectum

ICD code	Breast
C50.0	Malignant neoplasm of nipple and areola
C50.1	Malignant neoplasm of central portion of breast
C50.2	Malignant neoplasm of upper-inner quadrant of breast
C50.3	Malignant neoplasm of lower-inner quadrant of breast
C50.4	Malignant neoplasm of upper-outer quadrant of breast
C50.5	Malignant neoplasm of lower-outer quadrant of breast
C50.6	Malignant neoplasm of axillary tail of breast
C50.8	Malignant neoplasm of overlapping sites of breast
C50.9	Malignant neoplasm of breast of unspecified site

ICD code	Oesophagogastric
C15.0	Cervical part of oesophagus
C15.1	Thoracic part of oesophagus
C15.2	Abdominal part of oesophagus
C15.3	Malignant neoplasm of upper third of esophagus
C15.4	Malignant neoplasm of middle third of esophagus
C15.5	Malignant neoplasm of lower third of esophagus
C15.8	Malignant neoplasm of overlapping sites of esophagus
C15.9	Malignant neoplasm of esophagus, unspecified
C16.0	Malignant neoplasm of cardia
C16.1	Malignant neoplasm of fundus of stomach
C16.2	Malignant neoplasm of body of stomach
C16.3	Malignant neoplasm of pyloric antrum
C16.4	Malignant neoplasm of pylorus
C16.5	Malignant neoplasm of lesser curvature of stomach, unsp
C16.6	Malignant neoplasm of greater curvature of stomach, unsp
C16.8	Malignant neoplasm of overlapping sites of stomach
C16.9	Malignant neoplasm of stomach, unspecified

ICD code	Lung
----------	------

C33	Malignant neoplasm of trachea
C34.0	Malignant neoplasm of main bronchus
C34.1	Malignant neoplasm of upper lobe, bronchus or lung
C34.2	Malignant neoplasm of middle lobe, bronchus or lung
C34.3	Malignant neoplasm of lower lobe, bronchus or lung
C34.8	Malignant neoplasm of overlapping sites of bronchus and lung
C34.9	Malignant neoplasm of unspecified part of bronchus or lung

ICD code	Prostate
C61	Malignant neoplasm of prostate

iv Read codes for cancer diagnosis

Read code	Colorectal
BB5N100	[M]Adenocarcinoma in adenomatous polposis coli
BB5L100	[M]Adenocarcinoma in adenomatous polyp
BB5L300	[M]Adenocarcinoma in multiple adenomatous polyps
BB52000	[M]Adenocarcinoma in tubulovillous adenoma
BB5U100	[M]Adenocarcinoma in villous adenoma
BB52.00	[M]Adenocarcinoma NOS
BB57.00	[M]Adenocarcinoma, intestinal type
BB5y.00	[M]Adenoma and adenocarcinoms OS
BB5..00	[M]Adenomas and adenocarcinomas
BB5L.00	[M]Adenomatous and adenocarcinomatous polyps
BB5N.00	[M]Adenomatous and adenocarcinomatous polyps of colon
BB5Lz00	[M]Adenomatous or adenocarcinomatous polyp NOS
BB5Nz00	[M]Adenomatous or adenocarcinomatous polyps of the colon NOS
BB5Uz00	[M]Villous adenoma or adenocarcinoma NOS
BB5U.00	[M]Villous adenomas and adenocarcinomas
B1z0.11	Cancer of bowel
B141.11	Carcinoma of rectum
B13z.11	Colonic cancer
B139.00	Hereditary nonpolyposis colon cancer
B136.00	Malignant neoplasm of ascending colon
B134.00	Malignant neoplasm of caecum
B13..00	Malignant neoplasm of colon
B13z.00	Malignant neoplasm of colon nos
B132.00	Malignant neoplasm of descending colon
B130.00	Malignant neoplasm of hepatic flexure of colon
B18y200	Malignant neoplasm of mesorectum
B13y.00	Malignant neoplasm of other specified sites of colon
B140.00	Malignant neoplasm of rectosigmoid junction
B141.00	Malignant neoplasm of rectum

B14..00	Malignant neoplasm of rectum, rectosigmoid junction and anus
B133.00	Malignant neoplasm of sigmoid colon
B137.00	Malignant neoplasm of splenic flexure of colon
B131.00	Malignant neoplasm of transverse colon
B14z.00	Malignant neoplasm rectum,rectosigmoid junction and anus nos
B138.00	Malignant neoplasm, overlapping lesion of colon
B141.12	Rectal carcinoma
B134.11	Carcinoma of caecum

Read code	Lung
B2zy.00	Malignant neoplasm of other site of respiratory tract
BB5S212	[m]bronchiolar carcinoma
B561700	Secondary and unspec malig neop inferior tracheobronchial ln
B2...00	Malig neop of respiratory tract and intrathoracic organs
B223100	Malignant neoplasm of middle lobe of lung
B221000	Malignant neoplasm of carina of bronchus
B22y.00	Malignant neoplasm of other sites of bronchus or lung
B222z00	Malignant neoplasm of upper lobe, bronchus or lung nos
B551z00	Malignant neoplasm of thorax nos
B222100	Malignant neoplasm of upper lobe of lung
B22z.00	Malignant neoplasm of bronchus or lung nos
B222.00	Malignant neoplasm of upper lobe, bronchus or lung
B551.00	Malignant neoplasm of thorax
B223.00	Malignant neoplasm of middle lobe, bronchus or lung
BBLM.00	[m]pulmonary blastoma
B221.00	Malignant neoplasm of main bronchus
B224100	Malignant neoplasm of lower lobe of lung
BBTL.00	[m]intravascular bronchial alveolar tumour
B222.11	Pancoast's syndrome
B225.00	Malignant neoplasm of overlapping lesion of bronchus & lung
BB1K.00	[m]oat cell carcinoma
BB5S200	[m]bronchiolo-alveolar adenocarcinoma
B221z00	Malignant neoplasm of main bronchus nos
B22z.11	Lung cancer
B222000	Malignant neoplasm of upper lobe bronchus
B223000	Malignant neoplasm of middle lobe bronchus
B224000	Malignant neoplasm of lower lobe bronchus
B224.00	Malignant neoplasm of lower lobe, bronchus or lung
B22..00	Malignant neoplasm of trachea, bronchus and lung
B224z00	Malignant neoplasm of lower lobe, bronchus or lung nos
B223z00	Malignant neoplasm of middle lobe, bronchus or lung nos
Byu2000	[x]malignant neoplasm of bronchus or lung, unspecified

B221100	Malignant neoplasm of hilus of lung
---------	-------------------------------------

Read code	Breast
B34..00	Malignant neoplasm of female breast
B34..11	Ca female breast
B340.00	Malignant neoplasm of nipple and areola of female breast
B340000	Malignant neoplasm of nipple of female breast
B340100	Malignant neoplasm of areola of female breast
B340z00	Malignant neoplasm of nipple or areola of female breast nos
B341.00	Malignant neoplasm of central part of female breast
B342.00	Malignant neoplasm of upper-inner quadrant of female breast
B343.00	Malignant neoplasm of lower-inner quadrant of female breast
B344.00	Malignant neoplasm of upper-outer quadrant of female breast
B345.00	Malignant neoplasm of lower-outer quadrant of female breast
B346.00	Malignant neoplasm of axillary tail of female breast
B347.00	Malignant neoplasm, overlapping lesion of breast
B34y.00	Malignant neoplasm of other site of female breast
B34y000	Malignant neoplasm of ectopic site of female breast
B34yz00	Malignant neoplasm of other site of female breast NOS
B34z.00	Malignant neoplasm of female breast nos
BB90.00	[m]intraductal carcinoma, noninfiltrating nos
BB91.00	[m]infiltrating duct carcinoma
BB91100	[m]infiltrating duct and lobular carcinoma
BB94.00	[m]juvenile breast carcinoma
BB94.11	[m]secretory breast carcinoma
BB96.00	[m]noninfiltrating intraductal papillary adenocarcinoma
BB9F.00	[m]lobular carcinoma nos
BB9J.00	[m]paget's disease, mammary
BB9J.11	[m]paget's disease, breast
BB9K.00	[m]paget's disease and infiltrating breast duct carcinoma
BB9K000	[m]paget's disease and intraductal carcinoma of breast
Byu6.00	[X]Malignant neoplasm of breast

Read code	Stomach
B11..00	Malignant neoplasm of stomach
B11..11	Gastric neoplasm
B110.00	Malignant neoplasm of cardia of stomach
B110000	Malignant neoplasm of cardiac orifice of stomach
B110100	Malignant neoplasm of cardio-oesophageal junction of stomach
B110z00	Malignant neoplasm of cardia of stomach NOS
B111.00	Malignant neoplasm of pylorus of stomach
B111000	Malignant neoplasm of prepylorus of stomach

B111100	Malignant neoplasm of pyloric canal of stomach
B111z00	Malignant neoplasm of pylorus of stomach NOS
B112.00	Malignant neoplasm of pyloric antrum of stomach
B113.00	Malignant neoplasm of fundus of stomach
B114.00	Malignant neoplasm of body of stomach
B115.00	Malignant neoplasm of lesser curve of stomach unspecified
B116.00	Malignant neoplasm of greater curve of stomach unspecified
B117.00	Malignant neoplasm, overlapping lesion of stomach
B11y.00	Malignant neoplasm of other specified site of stomach
B11y000	Malignant neoplasm of anterior wall of stomach NEC
B11y100	Malignant neoplasm of posterior wall of stomach NEC
B11yz00	Malignant neoplasm of other specified site of stomach NOS
B11z.00	Malignant neoplasm of stomach NOS
BB55.00	[M]Linitis plastic

Read code	Prostate
4M0..00	Gleason grading of prostate cancer
4M00.00	Gleason prostate grade 2-4 (low)
4M01.00	Gleason prostate grade 5-7 (medium)
4M02.00	Gleason prostate grade 8-10 (high)
B46..00	Malignant neoplasm of prostate

v Read codes and terms for symptoms associated with colorectal cancer

Read code	Read term	Read code	Read term
Abdominal pain		Anaemia	
1962	colicky abdominal pain	1451	H/O: anaemia - iron deficient
1963	Non-colicky abdominal pain	1454	H/O: anaemia NOS
1964	Shoulder pain from abdomen	4235	Haemoglobin low
1967	type of git pain	6884	Anaemia screen
1968	abdominal discomfort	145..11	h/o: anaemia
1969	abdominal pain	2C2..00	o/e - anaemia
1971	central abdominal pain	2C2..11	o/e - anaemic
1972	Epigastric pain	2C22.00	O/E - equivocally anaemic
1977	right iliac fossa pain	2C23.00	O/E - clinically anaemic
1978	left iliac fossa pain	2C24.00	O/E - profoundly anaemic
1979	suprapubic pain	2C2Z.00	O/E - anaemia NOS
1969000	Abdominal wall pain	688..00	Anaemia/blood screening
196..00	Type of GIT pain	688..11	Anaemia screen
196..11	abdominal pain type	688Z.00	Anaemia/blood screen NOS
196..12	Type of GIT pain - symptom	B937X00	refractory anaemia, unspecified
196Z.00	Type of GIT pain NOS	C294300	Iron deficiency
197..00	Site of GIT pain	D0...00	deficiency anaemias
197..12	iliac fossa pain	D00..00	iron deficiency anaemias
197..13	site of abdominal pain	D00..11	hypochromic - microcytic anaemia
197A.00	generalised abdominal pain	D00..12	Microcytic - hypochromic anaemia
197A.11	General abdominal pain-symptom	D000.00	Iron deficiency anaemia due to chronic blood loss
197B.00	upper abdominal pain	D000.11	Normocytic anaemia due to chronic blood loss
197C.00	lower abdominal pain	D000.12	Iron deficiency anaemia due to blood loss
197D.00	Right upper quadrant pain	D001.00	Iron deficiency anaemia due to dietary causes
197Z.00	site of git pain	D00y.00	Other specified iron deficiency anaemia
25C..00	O/E - abdo. pain on palpation	D00y100	microcytic hypochromic anaemia
25C..11	O/E - epigastric pain on palp.	D00yz00	other specified iron deficiency anaemia nos
25C..14	O/E – umbilical pain on palpation	D00z.00	unspecified iron deficiency anaemia
25C..15	o/e - abdomen tender	D00z200	idiopathic hypochromic anaemia
25C2.00	O/E - abd.pain-R.hypochondrium	D00zz00	iron deficiency anaemia nos
25C3.00	O/E - abd. pain - epigastrium	D01..00	Other deficiency anaemias
25C4.00	O/E - abd.pain-L.hypochondrium	D01z.00	Other deficiency anaemias NOS
25C5.00	O/E - abd. pain - R.lumbar	D0y..00	Other specified deficiency anaemias
25C6.00	O/E - abd. pain - umbilical	D0z..00	deficiency anaemias nos
25C7.00	O/E - abd. pain - L.lumbar	D21..00	other and unspecified anaemias
25C8.00	O/E – abd. Pain - R iliac	D211.00	Acute posthaemorrhagic anaemia
25C9.00	O/e – abd. Pain – hypogastrium	D211.11	Normocytic anaemia following acute bleed
25CA.00	O/E – abd. Pain – L.ilic	D212.00	Anaemia in neoplastic disease
25CZ.00	O/E -abd.pain on palpation NOS	D213.00	Refractory Anaemia
25D2.00	O/E - guarding-R.hypochondrium	D214.00	Chronic anaemia
25D4.00	O/E - guarding-L.hypochondrium	D21y.00	Other specified anaemias
25E2.00	O/E - rebound-R.hypochondrium	D21yy00	Other specified other anaemia

25E4.00	O/E - rebound-L.hypochondrium	D21yz00	Other specified anaemia NOS
7NB2X00	[SO]Hypochondrium	D21z.00	anaemia unspecified
J574700	anal pain	D21z.11	Secondary anaemia NOS
J574800	rectal pain	D21z.12	normocytic anaemia due to unspecified cause
J574F00	Anorectal pain	D2y..00	Other specified anaemias
K58y000	Other pelvic pain - female	D2z..00	Other anaemias NOS
R073.00	[d]flatulence, eructation and gas pain	Dyu0.00	[X]Nutritional anaemias
R073200	gas pain (abdominal)	Dyu0000	[x]other iron deficiency anaemias
R073z00	[D]Flatulence, eructation and	Dyu2200	[X]Anaemia in other chronic diseases classified el
R079.00	[D] Defaecation painful	Dyu2400	[X]Other specified anaemias
R090.00	[d]abdominal pain	ZV78100	[V]Screening for other or unspecified deficiency a
R090000	[d]abdominal tenderness	Per rectal bleeding	
R090100	[d]abdominal colic	4762.11	Blood in faeces
R090200	[d]colic nos	14CA.11	H/O: GI Bleed
R090400	[d]abdominal cramps	196B.00	painful rectal bleeding
R090500	[D]Epigastric pain	196C.00	painless rectal bleeding
R090600	[D]Umbilical pain	19E6.00	Blood in faeces
R090700	[D] Hypochondrial pain	19E6.11	Blood in faeces symptom
R090800	[d]suprapubic pain	19ED.00	blood on toilet paper
R090900	[d]pain in right iliac fossa	19EG.00	Blood on pants
R090A00	pain in left iliac fossa	25T0.00	Bleeding stoma
R090E00	[d]recurrent acute abdominal p	G842000	Internal bleeding haemorrhoids
R090G00	[d]pelvic and perineal pain	G845000	External bleeding haemorrhoids
R090G12	[d] perineal pain	G848000	Bleeding haemorrhoids NOS
R090H00	[d]upper abdominal pain	J573.00	haemorrhage of rectum and anus
R090J00	[d]right upper quadrant pain	J573.11	bleeding pr
R090K00	[D]Left upper quadrant pain	J573000	rectal haemorrhage
R090L00	[D]Left lower quadrant pain	J573011	rectal bleeding
R090M00	[D]Right lower quadrant pain	J573012	prb - rectal bleeding
R090N00	nonspecific abdominal pain	J573z00	haemorrhage of rectum and anus nos
R090y00	[D]Other specified abdominal p	J681.11	blood in stool
R090z00	[d]abdominal pain nos	J681.12	Altered blood in stools
Ryu1100	[X]Other and unspecified abdom	J681.13	Blood in stools altered
Change in bowel habit		J68z.11	GIB - Gastrointestinal bleeding
19EA.00	change in bowel habit	J68z100	Intestinal haemorrhage NOS
19EA.11	altered bowel habit	Abdominal mass	
J52z.11	bowel dysfunction	2643	O/E – central pelvic mass
R078.00	[d]change in bowel habit	25J..00	o/e - abdominal mass palpated
X76d7	urgent desire for stool	25J2.00	O/E – abdominal mass < 1 quadrant
19C..00	constipation	25J3.00	O/E –abd.mass fills 1 quadrant
19C..11	constipation symptom	25J4.00	O/E – abdominal mass fills half abdomen
19C2.00	Constipated	25J5.00	O/E – abdominal mass fills abdomen
19CZ.00	Constipation NOS	25J7.00	right iliac fossa mass
19EC.00	Painful defaecation	25J8.00	O/E –left lower abdominal mass
2AF2.00	O/E - defaec.ref.abn.-constip.	25J9.00	Epigastric mass
E264500	Psychogenic constipation	25JZ.00	O/E – abdominal mass palpated NOS

J503100	faecal impaction	25K..00	O/E –abdominal mass consistency
J50zz12	Large bowel obstruction NOS	25K1.00	O/E – abdominal mass – soft
J520.00	constipation - functional	25K2.00	O/E – abdominal mass – hard
J520000	Acute constipation	25K3.00	O/E – abdominal mass very hard
J520100	chronic constipation with overflow	25KZ.00	O/E – abdominal mass consistency NOS
J520200	Chronic constipation without overflow	25L..00	O/E – abdominal mass shape
J520300	Drug induced constipation	25L1.00	O/E – abdominal mass – regular shape
J520400	Chronic constipation	25L2.00	O/E –abd.mass –irregular shape
J520y00	Other specified constipation	25LZ.00	O/E – abdominal mass – shape NOS
J520z00	constipation nos	25M..00	O/E – abdominal mass movt.with resp.
J52y100	Difficulty in ability to defaecate	25M1.00	O/E – abdominal mass moves with resp
Diarrhoea		25M2.00	O/E – abd.mass still with resp
19EE.00	Increased frequency of defaecation	25MZ.00	O/E – abdominal mass+ respn. NOS
19F..00	diarrhoea symptoms	25N..00	O/E – abdominal mass – border defined
19F..11	diarrhoea	25N1.00	O/E –abd.mass-upper border def
19F..12	loose stools	25N2.00	O/E –abd.mass –lower border def
19F2.00	diarrhoea	25NZ.00	O/E – abd.mass –border def. NOS
19F3.00	Spurious (overflow) diarrhoea	25Q3.00	o/e - pr - rectal mass
19FZ.00	Diarrhoea symptom NOS	25R3.00	O/e – dullness over abd. Mass
19FZ.11	diarrhoea & vomiting, symptom	264..00	o/e - pelvic mass palpated
19G..00	diarrhoea and vomiting	2I11.00	O/E - a lump
A076.11	Viral diarrhoea	R093.00	[d]swelling, mass or lump within abdomen or pelvis
A082.00	Infectious diarrhoea	R093000	[d]abdominal swelling
A082000	Dysenteric diarrhoea	R093100	[d]abdominal mass
A082z00	Infectious diarrhoea NOS	R093111	[D]Lump stomach
A083.00	Diarrhoea of presumed infectious origin	R093200	[d]abdominal lump
A083.11	Diarrhoea & vomiting -? infect	R093400	[d]pelvic mass
Ayu0H00	[X]Diarrhoea+gastroenteritis of presumed infectiou	R093700	[d]umbilical mass
E264300	Psychogenic diarrhoea	R093800	[D]Umbilical lump
E264311	Spurious diarrhoea	R093z00	[D] Swelling,mass orlump within abdomen/pelvis NOS
Eu45317	[X]Psychogenic diarrhoea	Fatigue or weight loss	
J4...13	Noninfective diarrhoea	1623	Weight decreasing
J432.11	Allergic diarrhoea	1625	abnormal weight loss
J433.11	Dietetic diarrhoea	1625.11	abnormal weight loss - symptom
J43z.11	chronic diarrhoea	1627	Unintentional weight loss
J4z..11	Presumed noninfectious diarrhoea	162..00	Weight symptom
J4zz.11	Diarrhoea - presumed non-infectious	162Z.00	Weight symptom NOS
J521000	Irritable bowel syndrome with diarrhoea	1D1A.00	complaining of weight loss
J525.00	Functional diarrhoea	22A6.00	O/E - Underweight
R077100	[d] stools loose	22A8.00	Weight loss from baseline weight
Other bowel function		22AZ.00	O/E - weight NOS
14C4.00	h/o: colitis	R032.00	[d]abnormal loss of weight
19A..00	Abdominal distension symptom	R034100	[D]Failure to gain weight

19B2.00	excessive flatulence	R034800	[D]Underweight
A081000	colitis - presumed infectious	1682	fatigue
J4...11	colitis - noninfective	1683	Tired all the time
J41..11	mucous colitis and/or proctiti	1683.11	c/o - "tired all the time"
J52..00	functional gastrointestinal tract disorders nec	168..00	tiredness symptom
J521.00	irritable colon - irritable bowel syndrome	168..11	fatigue - symptom
J521.11	irritable bowel syndrome	168..12	lethargy-symptom
R073000	[d]flatulence	168Z.00	Tiredness symptom NOS
R073300	[d]abdominal distension, gaseous	E205.12	tired all the time
R073z11	[d]wind	Eu46011	[X]Fatigue syndrome
		F286.00	Chronic fatigue syndrome
		F286.11	CFS - Chronic fatigue syndrome
		F286000	Mild chronic fatigue syndrome
		F286100	Moderate chronic fatigue syndrome
		F286200	Severe chronic fatigue syndrome
		R007.00	[D]Malaise and fatigue
		R007100	[d]fatigue
		R007300	[d]lethargy
		R007500	[d]tiredness
		R007z00	[D]Malaise and fatigue NOS

vi Read codes for referral for suspected colorectal cancer

Read code	Read term
8H48.00	Gastroenterological referral
8H5..11	Surgical referral
4794	Faecal occult blood: positive
8H51.00	General surgical referral
8HS..00	Referral for endoscopy
1J0..00	Suspected malignancy
ZL5AD00	Referral to gastroenterologist
ZL5G500	Referral to general surgeon
8HH8.00	Referred to cancer primary healthcare multidisciplinary team
8HU1.00	Referral for colonoscopy
ZL5GD00	Referral to colorectal surgeon
8HV0.00	Private referral to general surgeon
8H5J.00	Referral to colorectal surgeon
8HVN.00	Private referral to gastroenterologist
ZL54200	Referral to medical oncologist
8HS..11	Referral for gastroscopy
8H7o.00	Fast track referral
8H5Z.00	Referral to surgeon NOS
8H5..00	Referral to surgeon
8HU2.00	Referral for sigmoidoscopy
1J0D.00	Suspected upper gastrointestinal cancer
1J0E.00	Suspected lower gastrointestinal cancer
ZL5GA11	Referral to GI surgeon
8HHt.00	Fast track cancer referral
ZL5GE00	Referral to hepatobiliary surgeon
ZL5GA00	Referral to gastrointestinal surgeon
ZL5GC11	Referral to upper GI surgeon
ZL5GC00	Referral to upper gastrointestinal surgeon
ZL5G.00	Referral to surgeon
ZL5GB11	Referral to general GI surgeon
ZL5GB00	Referral to general gastrointestinal surgeon
8Hk5.00	Referred to hepatology service
ZL54.00	Referral to oncologist
8H5K.00	Referral to upper gastrointestinal surgeon
ZL62F00	Referral to oncology nurse
ZL5GF00	Referral to pancreatic surgeon
8Hn..00	Priority cancer referral

8Hn4.00	Fast track referral for suspected colorectal cancer
8Hn9.00	Fast track referral for suspected upper GI cancer
8HVc.00	Private referral to colorectal surgeon
8HkM.00	Referral to hepatobiliary and pancreatic surgery service
8Hm1.00	Referral to gastroenterology clinical assessment service
8H5P.00	Referral to gastrointestinal surgeon
8Hkn.00	Referral to community gastroenterology service
8HTE000	Referral to rectal bleeding clinic
8H5V.00	Referral to pancreatic surgeon
8HS0000	Referral to community sigmoidoscopy service
8HS0.00	Refer for sigmoidoscopy

vii OPCS codes for a colon cancer procedure

- H04-** Total excision of colon and rectum
- H05-** Total excision of colon
- H06-** Extended excision of right hemicolon
- H07-** Other excision of right hemicolon
- H08-** Excision of transverse colon
- H09-** Excision of left hemicolon
- H10-** Excision of sigmoid colon
- H11-** Other excision of colon
- H29-** Subtotal excision of colon
- G72.5** Anastomosis of ileum to anus and creation of pouch

viii OPCS code for chemotherapy

- X35.2** Intravenous chemotherapy
- X72.1** Delivery of complex chemotherapy for neoplasm including prolonged infusional treatment at first at
- X72.2** Delivery of complex parenteral chemotherapy for neoplasm at first attendance
- X72.3** Delivery of simple parenteral chemotherapy for neoplasm at first attendance
- X72.4** Delivery of subsequent element of cycle of chemotherapy for neoplasm
- X72.8** Other specified delivery of chemotherapy for neoplasm
- X72.9** Unspecified delivery of chemotherapy for neoplasm
- X73.1** Delivery of exclusively oral chemotherapy for neoplasm
- X73.8** Other specified delivery of oral chemotherapy for neoplasm
- X73.9** Unspecified delivery of oral chemotherapy for neoplasm
- X74.8** Other specified other chemotherapy drugs
- X74.9** Unspecified other chemotherapy drugs

ix ICD code for metastatic disease

- C78.0** Secondary malignant neoplasm of lung
- C78.1** Secondary malignant neoplasm of mediastinum
- C78.2** Secondary malignant neoplasm of pleura
- C78.3** Secondary malignant neoplasm of other and unspecified respiratory organs
- C78.4** Secondary malignant neoplasm of small intestine
- C78.5** Secondary malignant neoplasm of large intestine and rectum

- C78.6** Secondary malignant neoplasm of retroperitoneum and peritoneum
- C78.7** Secondary malignant neoplasm of liver and intrahepatic bile duct
- C78.8** Secondary malignant neoplasm of other and unspecified digestive organs
- C79** Secondary malignant neoplasm of other and unspecified sites
- C79.0** Secondary malignant neoplasm of kidney and renal pelvis
- C79.1** Secondary malignant neoplasm of bladder and other and unspecified urinary organs
- C79.2** Secondary malignant neoplasm of skin
- C79.3** Secondary malignant neoplasm of brain and cerebral meninges
- C79.4** Secondary malignant neoplasm of other and unspecified parts of nervous system
- C79.5** Secondary malignant neoplasm of bone and bone marrow
- C79.6** Secondary malignant neoplasm of ovary
- C79.7** Secondary malignant neoplasm of adrenal gland
- C79.8** Secondary malignant neoplasm of other specified sites
- C79.9** Secondary malignant neoplasm, unspecified site

x Reoperations

Wound operations including dehiscence

Re-enter abdomen for bleeding

Further resection

Relief of bowel obstruction inc. adhesiolysis or stent

Peritoneal washout

Formation or attention to stoma

Repair of organ inc perforation

Splenectomy

Repair of hernia

Perineal operation

Attention to anastomosis

Stent for fistula

Re-enter abdomen NEC

Operation on mesentery