

A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/152944>

Copyright and reuse:

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

The role of faecal calprotectin testing in primary care

by

Karoline Freeman

A thesis submitted in partial fulfilment of the requirements for the degree of
Doctor of Philosophy in Health Sciences

University of Warwick, Warwick Medical School

September 2020

Table of contents

Table of contents	2
List of figures	8
List of tables	11
List of boxes	14
Acknowledgments.....	15
Declaration	16
Abstract	17
List of abbreviations.....	18
Chapter 1 Introduction	19
1.1 My motivation for enquiry	19
1.2 Background.....	20
1.2.1 Abdominal symptoms in primary care	20
1.2.2 Overview of irritable bowel syndrome and inflammatory bowel disease ..	20
1.2.3 Testing patients with abdominal symptoms.....	27
1.2.4 Lack of evidence from primary care – the rationale for this thesis	29
1.3 Aim and Objectives	31
1.4 Thesis outline	31
Chapter 2 Choosing an approach	36
2.1 Chapter overview	36
2.2 Research paradigm or world view	36
2.3 Framework approach for qualitative data.....	37
2.3.1 Inductive-deductive balance.....	38
2.3.2 Validity and generalisability	38
2.4 Methodologies for quantitative data.....	39
2.4.1 Systematic review	39
2.4.2 Retrospective cohort study.....	39
2.4.3 Validity and generalisability	40

2.5 Summarising my overall research approach.....	40
Chapter 3 Use of FC testing in primary care	43
3.1 Chapter overview	43
3.2 Introduction	43
3.2.1 Routine data for primary care research.....	43
3.2.2 THIN data to address the study objectives.....	46
3.3 Objectives	47
3.4 Methods	47
3.4.1 Design of the THIN study.....	47
3.4.2 Methods of creating THIN datasets.....	48
3.4.3 Methods of analysis	53
3.4.4 Summarising the methods	56
3.5 Results	56
3.5.1 IBD prevalence and incidence	57
3.5.2 FC test uptake	60
3.5.3 Characterisation of FC test use.....	63
3.5.4 Referral and colonoscopy following FC testing.....	69
3.6 Discussion.....	74
3.6.1 Summary of study findings.....	74
3.6.2 Strengths and limitations of primary care electronic health care records for research	74
3.6.3 Findings in context of published evidence.....	78
3.7 Summarising the use of FC testing in primary care	81
Chapter 4 Test accuracy of faecal calprotectin testing for IBD	83
4.1 Chapter overview	83
4.2 Systematic review and exploratory meta-analysis of diagnostic test accuracy studies of FC testing	84
4.2.1 Introduction.....	84
4.2.2 Objectives.....	85
4.2.3 Methods.....	85

4.2.4 Results.....	88
4.2.5 Discussion	98
4.3 Test accuracy of FC testing in primary care according to THIN	101
4.3.1 Introduction.....	101
4.3.2 Objectives.....	102
4.3.3 Methods.....	102
4.3.4 Results.....	109
4.3.5 Discussion	128
4.4 Tailored meta-analysis of published test accuracy studies	136
4.4.1 Introduction.....	136
4.4.2 Objectives.....	137
4.4.3 Methods.....	137
4.4.4 Results.....	139
4.4.5 Discussion	141
4.5 Summarising the test accuracy of FC testing.....	144
Chapter 5 Influences on GPs' referral decision and the effects of FC testing.....	147
5.1 Chapter overview	147
5.2 Introduction	147
5.2.1 Rationale	147
5.2.2 Background	148
5.2.3 Theoretical framework	148
5.3 Objectives	149
5.4 Methods	150
5.4.1 Study design.....	150
5.4.2 Ethical considerations	150
5.4.3 Sample and recruitment.....	150
5.4.4 Data collection	151
5.4.5 Analysis	152
5.4.6 Patient & Public Involvement	153

5.5 Results	154
5.5.1 Influences on referral decisions.....	158
5.5.2 FC testing within the referral decision	177
5.6 Discussion.....	192
5.6.1 Main empirical findings on referral decisions.....	192
5.6.2 Theoretical findings in the context of published models of referral decisions	196
5.6.3 FC testing as a referral decision aid.....	200
5.6.4 Strengths and limitations.....	202
5.6.5 Validation of study findings	203
5.6.6 Reflections.....	204
5.7 Summarising the main influences on referral decisions and the impact of FC testing on referral decisions.....	205
Chapter 6 Impact of study findings on cost-effectiveness outcomes in the NICE assessment	209
6.1 Chapter overview	209
6.2 Health economic model.....	210
6.2.1 Model structure and assumptions	210
6.2.1 Model base case parameters and outcomes.....	211
6.2.3 Sensitivity and scenario analyses	212
6.3 Potential impact of findings from Chapters 3, 4, and 5 on cost-effectiveness findings	213
6.3.1 Discussing model outcomes considering test accuracy based on My THIN data	213
6.3.2 Discussing model outcomes considering clinician behaviour	216
6.4 Summarising the impact of my study findings on the cost-effectiveness of FC testing	216
Chapter 7 Discussion.....	219
7.1 Chapter overview	219
7.2 Meeting my study objectives	219

7.2.1 Summarising my main aim and objectives	219
7.2.2 Summarising my contribution to research	219
7.2.3 Test accuracy of FC testing in primary care in light of NICE guidance .	221
7.2.4 FC test use in primary care in light of NICE guidance	226
7.2.5 Generalisability of study findings.....	227
7.3 Implications for clinical practice	228
7.4 Implications for future research	229
Chapter 8 Conclusions.....	232
References	235
Appendix	247
Appendix A Use of FC testing in primary care	247
A1 Quality indicators of THIN data for research purposes.....	247
A2 Brief overview of Read codes	247
A3 Creating code lists for study variables	248
A4 Exploration of the THIN dataset and management decisions of irregular cases	268
Appendix B Test accuracy of FC testing for IBD.....	270
B1 Search strategy	270
B2 Between study variance of logit sensitivity and logit specificity of each random effects meta-analysis (for tests and clinical questions).....	272
B3 R code for my exploratory 25,000 meta-analyses and 2D density plot.....	273
B4 Excluded studies with reason	275
B5 Tables of study characteristics	288
B6 Risk of bias and applicability concern for each included study.....	304
B7 Paired forest plots of sensitivity and specificity of 38 included studies by clinical question	305
B8 2x2 table data of studies included in meta-analysis by test (50µg/g threshold)	307
B9 2x2 table data of studies included in meta-analysis by clinical question (50µg/g threshold)	308

B10 2x2 table data at 100µg/g threshold by clinical question	309
B11 2x2 table data at 50µg/g threshold of 28 studies considered for exploratory sensitivity analysis	310
B12 Setting criteria for inclusion of patients into the test accuracy study.....	312
B13 Description of THIN variables relating information on the FC test result	318
B14 Association of variables with type of diagnosis by FC test outcome.....	319
B15 Defining the applicable region and selecting studies for tailored meta-analysis	321
B16 R code to select studies for tailored meta-analysis	323
Appendix C Influences on GPs' referral decision and the effects of FC testing.	329
C1 Indicative interview schedule	329
C2 Coding framework for indexing interview transcripts.....	331
C3 Categorisation of coded elements into higher order categories and subthemes for the doctor-patient interaction theme	332

List of figures

Figure 1.1 Visual outline of thesis structure	34
Figure 2.1 Logic model depicting my approach of investigation	41
Figure 3.1 Hypothetical THIN study cohort of individual patients entering and leaving the study within the study period resulting in different length of follow-up times	48
Figure 3.2 Flow diagram of inclusion and exclusion of patients into the analyses of IBD prevalence and incidence	57
Figure 3.3 12-month period prevalence of IBD per 10,000 adult THIN population	58
Figure 3.4 Annual IBD incidence per 100,000 adult THIN population	60
Figure 3.5 Overview of inclusion and exclusion of patient records with and without FC tests into the study of FC uptake	61
Figure 3.6 FC tested patients per 1000 THIN population by year	61
Figure 3.7 Number of practices with first ever FC test record by year	62
Figure 3.8 Number of FC tests recorded per 1000 practice population by number of practices during 2006-2016	62
Figure 3.9 Overview of inclusion and exclusion to create a dataset of patients with eligible symptoms, FC tests and/or IBD diagnosis	66
Figure 3.10 Number of patients with eligible symptoms for FC testing, a record of FC testing and/or an IBD diagnosis for four year period and for 2016 only	67
Figure 3.11 Time from FC testing to referral by FC test result (the dotted line marks six weeks following FC testing)	69
Figure 3.12 Time from FC testing to colonoscopy by FC test result (the dotted line marks one year following FC testing)	70
Figure 3.13 Patients with sufficient follow-up available after FC testing for consideration of referral within six months and colonoscopy within one year following FC testing	71
Figure 3.14 Proportions of IBD diagnoses in referred and colonoscoped patients by FC test result	72
Figure 4.1 Flow diagram of study inclusion	89
Figure 4.2 Overview of risk of bias and applicability concern of included studies according to QUADAS-2	91
Figure 4.3 Forrest plot of A) Oyaert 2017 ¹²⁵ comparing 6 FC tests and B) Labaere 2014 ¹²⁶ comparing five FC tests at cut-off 50µg/g	92

Figure 4.4 Receiver operating characteristic (ROC) plot of sensitivity and specificity by test	94
Figure 4.5 Receiver operating characteristic (ROC) plot of sensitivity and specificity by clinical question	94
Figure 4.6 Pairs of sensitivity and specificity of 25,000 meta-analyses of 28 studies picking one outcome per study at random for each round of meta-analysis at a threshold of 50µg/g.	97
Figure 4.7 Flow diagram of inclusion criteria of FC tests into test accuracy study	112
Figure 4.8 Overview of diagnoses prior and following FC testing for 7,084 included FC tested patients	114
Figure 4.9 Datasets of FC tested patients with three different follow-up times available for three analyses	114
Figure 4.10 ROC curve of sensitivity and false positive rate (1-specificity) for thresholds 33-300µg/g for the differential IBD versus non-IBD for tests with at least six months (N=5,970), 12 months (N=4,793) and 24 months (N=2,662) follow-up available after the FC test date	116
Figure 4.11 Datasets for three follow-up times of patients with numeric FC test results preceded by “=”	117
Figure 4.12 ROC curve of sensitivity and false positive rate (1-specificity) for all thresholds considering test results reported with operator “=” for the differential IBD versus non-IBD for patients with at least six months (N=3,836), 12 months (N=3,053) and 24 months (N=1,722) follow-up available after the FC test date	117
Figure 4.13 Sensitivity and specificity at FC thresholds 33-300µg/g for IBD versus non-IBD for patients with six months follow-up (N=5,970 any operator)	118
Figure 4.14 Illustration of the size of the available datasets for the comparison of test accuracy for the two differentials IBD versus non-IBD and IBD versus IBS	119
Figure 4.15 Comparison of ROC curves of FC testing for IBD versus non-IBD (n=3,836) and IBD versus IBS (n=304) considering FC test results recorded with the “=” operator	120
Figure 4.16 Proportions of patients with and without IBD/IBS/CRC diagnosis following FC testing by test outcome	121
Figure 4.17 Kaplan Meier time to event curve depicting time to diagnosis for cases of IBD, IBS and CRC from the time of FC testing	122
Figure 4.18 ROC plot of studies reporting the sensitivity and specificity of FC testing for IBD at the threshold of 50µg/g	140
Figure 4.19 Pairs of sensitivity and false positive rate from conventional meta-analysis, tailored meta-analysis and THIN data in ROC space and in relation to the applicable region informed by routine data from THIN	141

Figure 5.1 Map of recruitment from the seven Clinical Commissioning Groups (CCGs) of the West Midlands South region; depicting the number of GPs recruited from each CCG	154
Figure 5.2 Overview of four influences (green ovals) on referral decisions and outcomes of FC testing	158
Figure 5.3 Theme 1 of referral influences - clinical need - and its subthemes	158
Figure 5.4 Theme 2 of referral influences - Doctor-patient interaction - and its subthemes	162
Figure 5.5 Theme 3 of referral influences - GP factors - and its subthemes	166
Figure 5.6 Theme 4 of referral influences - Health care system - and its subthemes	172
Figure 5.7 Overview of referral influences highlighting the outcomes of FC testing and the direction of their impact on the referral decision	178
Figure 5.8 Overview of outcomes of FC testing and their effect on the four influences of referral decisions	179
Figure 5.9 GPs' interpretation of the role of FC testing in general practice	183
Figure 5.10 GPs' views on the place of FC testing in general practice	186
Figure 5.11 Requirements for FC testing according to GPs	188
Figure B1 Forest plots of 14 studies reporting data for clinical questions IBD vs IBS and IBD vs functional disease	305
Figure B2 Forest plots of studies reporting data for clinical questions IBD vs non-IBD and Inflammatory vs non-inflammatory intestinal disease	306
Figure B3 Forest plots of studies reporting data for clinical questions organic intestinal disease vs non-organic intestinal disease and organic intestinal disease vs other intestinal conditions	306
Figure B4 Distribution of numeric FC test results recorded in A) $\mu\text{g/g}$, B) mg/kg and C) mmol/l	314
Figure B5 Distribution of numeric FC results in $\mu\text{g/g}$ preceded by the ">" operator	315
Figure B6 Distribution of numeric FC results in $\mu\text{g/g}$ preceded by the "<" operator	316
Figure B7 Distribution of numeric FC results in $\mu\text{g/g}$ preceded by the "=" operator	317
Figure B8 Hypothetical example explaining the applicable region in ROC space	322

List of tables

Table 3.1 Definition of conditions for the THIN study	50
Table 3.2 Definitions of symptoms for the THIN study	51
Table 3.3 Definitions of therapeutic drugs for the THIN study	52
Table 3.4 12-month period prevalence of IBD per 10,000 adult THIN population	58
Table 3.5 Annual IBD incidence per 100,000 for the adult THIN population	59
Table 3.6 Number of tests per patient between 2009 and 2016	63
Table 3.7 Demographic and clinical characteristics of FC tested patients per recorded FC test between 2009 and 2016	65
Table 3.8 IBD prevalence in eligible patients for FC testing defined as patients with 1) a recorded period of NICE eligible symptoms and 2) a recorded FC test for the study period 2013-2016	68
Table 3.9 Proportions of referred and colonoscoped FC tested patients in published studies compared to the present study of FC tested patients with at least 1 year follow-up	73
Table 4.1 Summary of study characteristics of included studies addressing the test accuracy question of FC testing in the detection of IBD	89
Table 4.2 Comparison of results of separate meta-analyses and meta-regression with test type and clinical question as covariates at two thresholds	95
Table 4.3 Information available from THIN variables for 17,466 FC tests recorded in THIN relating to the FC test result	110
Table 4.4 Patient characteristics of FC tested patients with appropriate test result (included) and inappropriate test result (excluded)	113
Table 4.5 Test accuracy measures of FC testing at threshold of 50µg/g	115
Table 4.6 Test accuracy measures of FC testing at threshold of 100µg/g compared to 50µg/g	119
Table 4.7 Test accuracy measures for FC testing at the threshold of 50µg/g for the detection of IBD recorded within six months of FC testing considering two different non-target conditions	120
Table 4.8 Sensitivity analyses compared to primary analysis of FC testing for IBD vs non-IBD at a threshold of 50µg/g with IBD recorded within six months	124
Table 4.9 Patients referred by test result and disease status	125

Table 4.10 Estimates of sensitivity and specificity for FC testing for IBD versus non-IBD based on the THIN analysis compared with published results from primary care and meta-analytical results of primary and secondary care studies at two thresholds	127
Table 4.11 Comparison of sensitivity and specificity at the common threshold of 50µg/g from tailored meta-analysis with conventional meta-analysis and results using THIN data	141
Table 5.1 Characteristics of participating general practices	155
Table 5.2 Characteristics of participating GPs	156
Table 5.3 Overview of themes and subthemes describing the influences on the referral decision	157
Table 6.1 Estimates of sensitivity and specificity considered in the cost-utility model and estimates based on my THIN data study	211
Table 6.2 Results from base case and scenario analysis considering a wider testing population	212
Table 6.3 Disease prevalence and proportions of true positives, false negatives, true negatives and false positives in the detection of IBD by GP assessment versus FC testing for base case assumptions and the scenario analysis of a wider testing population based on sensitivity and specificity estimates	212
Table 6.4 Main model parameters for base case and scenario analysis compared to findings from THIN analysis	214
Table 6.5 Possible impact of changing from the original model assumptions to my study outcomes on cost-effectiveness findings	215
Table 7.1 Impact of mechanisms identified by Leeflang (2009) on the patient population and FC test accuracy in my primary care study population compared to secondary care studies	222
Table A1 Final code lists for study variables	250
Table A2 Frequencies of IBD records identified as illogical or inaccurate	268
Table B1 Between study variance of logit sensitivity and logit specificity for tests and clinical questions	272
Table B2 Full texts excluded from review	275
Table B3 Abstracts excluded from review	281
Table B4 Studies excluded from previous review	286
Table B5 Study characteristics - population	288
Table B6 Study characteristics – investigations	297

Table B7 risk of bias and applicability concern for each included study	304
Table B8 2x2 table data of studies included in meta-analysis - EK-CAL	307
Table B9 2x2 table data of studies included in meta-analysis - ELiA	307
Table B10 2x2 table data of studies included in meta-analysis - PhiCal	307
Table B11 2x2 table data of studies included in meta-analysis - Quantum-Blue	307
Table B12 2X2 table data of studies included in meta-analysis - IBD versus IBS	308
Table B13 2X2 table data of studies included in meta-analysis - IBD versus non-IBD	308
Table B14 2X2 table data of studies included in meta-analysis - Organic versus non-organic	309
Table B15 2X2 table data of studies included in meta-analysis (100µg/g threshold) - IBD versus IBS	309
Table B16 2X2 table data of studies included in meta-analysis (100µg/g threshold) -IBD versus non-IBD	309
Table B17 2X2 table data of studies included in meta-analysis (100µg/g threshold) - Organic versus non-organic	310
Table B18 28 Studies considered for exploratory sensitivity analysis	310
Table B19 THIN variables relating information on the FC test result	313
Table B20 Common test statistics for distributions of numeric FC test results reported as µg/g, mg/kg and mmol/l	316
Table B21 Summary statistics of the distribution of numeric FC results in µg/g preceded by the ">" operator	316
Table B22 Summary statistics of the distribution of numeric FC results in µg/g preceded by the "<" operator	317
Table B23 Summary statistics of the distribution of numeric FC results in µg/g preceded by the "=" operator	318
Table B24 Association of symptoms, referral, further testing and comorbidities with type of diagnosis by FC test outcome based on 2,414 positives and 3,556 negative FC tests	319
Table B25 Contingency table explaining the mathematic relationship between the test positive rate, the prevalence, the sensitivity and the false positive rate	321
Table C1 Categorisation of coded elements into higher order categories and subthemes for the doctor-patient interaction theme	332

List of boxes

Box 1.1 Definition of IBS according to Rome IV criteria in Lacy (2017)	22
Box A1 Definitions of THIN quality indicators	247
Box A2 Creating Read code lists for IBD	249

Acknowledgments

First of all, I would like to acknowledge the NIHR for funding my research and training¹. I would like to say a special thank you to my supervisors Professor Aileen Clarke, Professor Sian Taylor-Phillips and Dr Brian Willis for their guidance and encouragement, their constructive feedback and thoughtful questions and their kindness and reassurance. I would also like to thank my advisory group, in particular Professor Kate Seers, Dr Margaret Glogowska and Dr Ronan Ryan, for their time and helpful input and Dr Nick Parsons for his statistical support. Further, I would like to thank my PPI advisors for their support and interest in my research, the GPs who took part in my study for their willingness to be interviewed and the Primary Care Clinical Research Network for helping me with the recruitment of GP participants. My thanks also go to the Test Evaluation Research Group at Birmingham University and the NIHR MedTech and In Vitro Diagnostics Co-operative at Oxford University for hosting my research visits and for insightful discussions. I am also thankful to colleagues and friends for their encouragement and listening ears and their help in proof reading my main chapters. Finally, I would like to say thank you to my family for their patience and support over the four years.

¹ Karoline Freeman, Doctoral Research Fellow, DRF-2016-09-038, is funded by the National Institute for Health Research (NIHR) for this research project. The views expressed in this publication are those of the author and not necessarily those of the NIHR, NHS or the UK Department of Health and Social Care.

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

The work presented (including data generated and data analysis) was carried out by the author except in the cases outlined below:

Chapters 3 and 4.3 are based on routine electronic health care records from the Health Improvement Network (THIN) database. I produced code lists and definitions for the study variables. A specialised THIN data analyst used the code lists to extract and collate patient data into a dataset. The dataset was returned to me for cleaning, management and analysis.

Parts of this thesis have been published by the author:

Published:

Freeman K, Willis BH, Fraser H, Taylor-Phillips S and Clarke A. *Faecal calprotectin to detect inflammatory bowel disease: a systematic review and exploratory meta-analysis of test accuracy*. BMJ Open 2019; 9:e027428.

Submitted:

Freeman K, Taylor-Phillips S, Willis BH, Ryan R and Clarke A. *Test accuracy of faecal calprotectin for inflammatory bowel disease in UK primary care: retrospective cohort study of the IMRD-UK data*. Submitted to BMJ Open on 25/08/2020

Abstract

Background: Faecal calprotectin (FC) testing has received national approval for the detection of inflammatory bowel disease (IBD) in primary care. Approval was based on evidence taking the secondary care perspective. Transferring evidence between settings may not be appropriate.

Objectives: To produce primary care evidence on the test accuracy, test use and patient management following FC testing. To identify biases and misconceptions in the current evidence base of FC testing. To assess the appropriateness of national recommendations on testing in primary care.

Methods: Retrospective cohort study of primary care electronic health records to investigate the uptake and use of FC testing in routine practice. Triangulation of test accuracy measures using routine primary care data and tailored meta-analysis of studies deemed applicable to primary care. Semi-structured interviews with general practitioners (GPs) to explore influences on referral decisions and the perceived role of FC testing in referral decisions.

Results: Sensitivity of FC testing was high and similar across methods and settings. Specificity was more variable mainly due to differences in the definition of the non-IBD group. Test use was variable and inconsistent. Few patients followed the anticipated patient pathway. FC test negative referrals were close to 50%. Qualitative findings suggested that reasons included clinical uncertainty and patient preferences. Factors other than FC test results were more influential on referral decisions. Secondary care test accuracy estimates and assumptions on test use in the cost-effectiveness model were not applicable to the clinical problem in primary care.

Conclusions: FC testing is a useful test for primary care. It may not be cost-effective when used in primary care currently because the intended use does not match the GPs' reasons for testing and referral. Test evaluations and national guidance for primary care need to consider the primary care context and GPs' decision-making processes for testing and referral decisions.

List of abbreviations

AHD	Additional health data
AMR	Acceptable mortality reporting date
BMI	Body mass index
BNF	British National Formulary
CCG	Clinical commissioning group
CD	Crohn's disease
CI	Confidence interval
CPRD	Clinical practice research datalink
CRC	Colorectal cancer
CRN	Clinical research network
DG	Diagnostics guidance
FC	Faecal calprotectin
FN	False negative
FP	False positive
GP	General practitioner
HES	Hospital episodic statistics
IBD	Inflammatory bowel disease
IBS	Irritable bowel syndrome
IQR	Interquartile range
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
POCT	Point of care test
PPI	Patient and public involvement
QALY	Quality adjusted life year
QoL	Quality of life
QUADAS	Quality Assessment of Diagnostic Accuracy Studies
ROC	Receiver operating characteristic
SD	Standard deviation
THIN	The health improvement network
TN	True negative
TP	True positive
UC	Ulcerative colitis

Chapter 1 Introduction

1.1 My motivation for enquiry

Prior to starting my PhD, I worked as a systematic reviewer producing reviews and health technology assessments for the National Institute for Health and Care Excellence (NICE) and the National Screening Committee. Over the years I have been involved in a number of test evaluations covering different types of tests for screening, monitoring, diagnosis and prognosis and covering a range of conditions. These assessments were based on limited published evidence which was often of poor quality. Nevertheless, national decision makers formulated decisions about the availability of these tests for patient care. I was interested to explore what level of evidence is sufficient to support national decisions and when the assessment of limited evidence may lead to wrong decisions.

I explored the evidence base of an approved test, namely faecal calprotectin (FC) for the detection of inflammatory bowel disease (IBD). The test was approved by NICE for use in primary care without any evidence from the primary care setting in terms of its test accuracy and impact on subsequent management decisions. The NICE assessment assumed that evidence from other settings could be transferred and that test use would follow the outcomes of studies of test accuracy, costs and benefits. There appeared to be a need to investigate the assumptions and to inform decision makers about possible consequences when using such an approach. In this thesis, I report my investigations into the role of faecal calprotectin testing in patients with abdominal symptoms presenting to their general practitioners (GP) in primary care.

1.2 Background

1.2.1 ABDOMINAL SYMPTOMS IN PRIMARY CARE

"I always joke with colleagues and say the one thing that scares me is abdominal pain when they come in, because it could be anything" (GP19)

Abdominal symptoms present a diagnostic challenge to GPs and often involve long consultations, challenging patients and repeat visits. In the UK about 25% of the population have symptoms of abdominal pain or discomfort at any one time, and approximately 2.5% of the UK population consult a GP each year because of abdominal pain.¹ As a result gastrointestinal symptoms are common in primary care accounting for about one in 12 consultations.² In a considerable proportion of patients the underlying cause cannot be determined. In those patients with a diagnosis, irritable bowel syndrome (IBS) – a functional disorder - is the second most common aetiology of patients with abdominal pain after gastroenteritis, an infection of the bowel. About 1 in 10 patients suffer from a serious organic condition.³ It is the GP's role to differentiate between patients who need to be referred for further investigations and those who can be managed in primary care. In adults, following the exclusion of infectious gastroenteritis, coeliac disease and more serious conditions such as cancer based on history, examination and first line tests, the need to refer is commonly based on distinguishing between patients with severe IBS and IBD. In practice this is often challenging as symptoms of IBS and IBD overlap.

In the following I present an overview of the two conditions, IBS and IBD, introduce the FC test and provide the rationale for my investigations. I then list my study objectives and provide a thesis outline.

1.2.2 OVERVIEW OF IRRITABLE BOWEL SYNDROME AND INFLAMMATORY BOWEL DISEASE

1.2.2.1 Irritable bowel syndrome

1.2.2.1.1 Epidemiology – the distribution of disease

IBS is a common, chronic condition of the gastrointestinal tract characterised by the presence of abdominal pain with constipation, diarrhoea or both. IBS symptoms are present in between 3-22% of the general western population.⁴ This variation in prevalence estimates is due to differences in study populations and diagnostic criteria used to identify IBS. IBS is roughly twice as common in women as in men, it

is more prevalent in young adults and twice as prevalent in urban areas compared to rural areas.⁵ It is estimated that between 20-30% of these symptomatic people in the general population seek GP advice and become IBS patients resulting in an IBS prevalence of about 2.5% in primary care,² and approximately 79,000 new IBS diagnoses every year.⁶ About 20% of primary care IBS patients are referred to gastroenterology.²

It is generally assumed that patients managed in primary care have less severe symptoms, a shorter history of symptoms and fewer psychological problems than those seen in specialist care. Furthermore, people with IBS symptoms in the general population are different to IBS patients seeking health care.⁵ However, most of our knowledge of IBS comes from studies of patients referred to gastroenterology clinics⁷ and might, therefore, only be applicable to a select subgroup of patients.

1.2.2.1.2 Natural history

The natural history of IBS is difficult to ascertain because natural history studies in IBS are often affected by treatments introduced by the patient or clinician.⁸ IBS is a chronic condition characterised by intermittent flares, but we have little knowledge about the frequency and duration of flares.⁷ Hahn et al. (1998) reported in an early study that patients experienced a mean of 12 distinct episodes of active IBS symptoms during a 12-week period lasting each for a maximum of five days.⁹ While the type of symptom experienced by individual patients remains relatively stable,¹⁰ the severity may vary over time. Over a 2-year period between 12-38% of IBS patients experienced complete disappearance of symptoms, 30-50% had unchanged symptoms and symptoms worsened for 2-18% of patients.⁷

IBS patients often present with at least one co-morbidity contributing to the disease burden. Common functional comorbidities include somatic pain syndrome, functional dyspepsia and depression.^{7, 11} The likelihood of developing an organic gastrointestinal disorder long term is low at 2-5%.⁷ IBS is not associated with increased mortality.¹²

1.2.2.1.3 Aetiology and pathophysiology - underlying causes and effects

The pathophysiology of IBS is not definitely known. IBS is believed to be caused by a multitude of factors including genetics, diet, stress, infection, and medicines such as antibiotics.⁸ Through different mechanisms these lead to the typical symptoms of IBS, namely, abnormal gastrointestinal function and motility. The mechanisms

include altered pain perception, altered brain-gut interaction, microbial imbalance, increased intestinal permeability, increased gut mucosal immune activation and visceral hypersensitivity. No single trigger could be identified to be leading in the cause of IBS. It is, therefore, believed that IBS consists of a number of different conditions of different causes with similar symptoms.⁸

1.2.2.1.4 Diagnosis

To date no precise biomarker for IBS exists, therefore, a definitive test for IBS is not available. Diagnosis of IBS is based on the presence of characteristic symptoms for which a number of different criteria have been developed over the years including those by Manning,¹³ Kruis¹⁴ and a group of international experts (Rome criteria).¹⁵ The Rome criteria were a response to shortcomings of the earlier criteria by Manning and Kruis. The third iteration of the Rome criteria achieved sensitivity and specificity of 68.8% and 79.5%, respectively.¹⁶ Changes to Rome III aimed to make the criteria more clinically useful.¹⁵ The resulting Rome IV criteria define IBS in terms of type and duration of symptoms (Box 1.1). However, there is general belief that the available criteria are for research and are not usually used in clinical practice to diagnose IBS.¹⁷ In current general practice IBS is a diagnosis of exclusion of alternative organic conditions and most patients undergo extensive diagnostic evaluations before, both, patients and clinicians accept an IBS diagnosis.⁷ The current NICE guidelines CG61 recommend an initial assessment of symptoms including abdominal pain or discomfort, bloating and change in bowel habit and exclusion of red flags for bowel cancer and severe IBD. This should be followed by investigations to exclude other conditions such as coeliac disease before an IBS diagnosis is concluded.¹⁸

Box 1.1 Definition of IBS according to Rome IV criteria in Lacy (2017)¹⁵

“IBS is a functional bowel disorder in which recurrent abdominal pain is associated with defecation or a change in bowel habits. Disordered bowel habits are typically present (i.e., constipation, diarrhea or a mix of constipation and diarrhea), as are symptoms of abdominal bloating/distension. Symptom onset should occur at least six months prior to diagnosis and symptoms should be present during the last three months.”

1.2.2.1.5 Management

The vast majority of IBS patients are managed in primary care with some patients not requiring any medication.¹² The management of global IBS symptoms requires a

holistic approach by a trusted physician and includes an explanation of the condition and an explanation of the importance of self-management. NICE clinical guidelines recommend the provision of information on general lifestyle, physical activity and diet followed by symptom targeted medication.¹⁸ Medication should be based on the severity and nature of symptoms, the type and degree of possible psychological disorders and the degree of functional impairment.⁵ There is no specific or standardized treatment and many of the medications are available over the counter. Symptoms of diarrhoea and constipation are generally treated with antidiarrhoeal agents, antispasmodics including peppermint oil, fibre supplements and laxative agents. Probiotics and antibiotics can be trialled to modify the gut microbiota while antidepressants are widely used because of their effects on pain perception, mood and motility.⁸ Psychological therapies including behavioural therapy should be considered for people who do not respond to IBS medications.¹⁸

1.2.2.2 Inflammatory bowel disease

1.2.2.2.1 Epidemiology – the distribution of disease

IBD is a progressive, organic disease characterised by inflammation of the intestine causing bowel damage and disability.^{19, 20} The term describes two main chronic relapsing disorders called ulcerative colitis (UC) and Crohn's disease (CD). Both are typically characterised by abdominal pain, bloody diarrhoea and weight loss. UC only affects the large intestine while CD can affect any part of the digestive system. Indeterminate colitis is diagnosed when it is unclear whether the symptoms represent CD or UC. Additionally, microscopic colitis is a less common type of IBD where inflammation is only visible under a microscope. IBD affects over 250,000 patients in the UK,²¹ which translates into a prevalence of between 0.3-0.4%.²²⁻²⁵ There are approximately 10,000 new IBD diagnoses every year.²¹ Incidence and prevalence of IBD are increasing and are greater in urban areas than in rural areas.^{19, 20} Men and women are affected equally. Disease onset peaks in early adulthood with a second peak between 50 and 60 years of age for Crohn's disease.²⁰

1.2.2.2.2 Natural history

Crohn's disease and ulcerative colitis are progressive, relapsing and remitting diseases.^{19, 20} Both conditions have an unpredictable clinical course but they differ significantly in presentation and progression.

CD can affect the whole gastrointestinal tract from the mouth to the anus and inflammation affects the full thickness of the bowel wall. Complications, such as strictures, fistula and abscesses, lead to surgery in half of the patients within 10 years.^{20, 26} Surgery is not curative with 50% of patients having recurrence of clinical features and 80% having recurrence of inflammation on endoscopy.²⁰

In UC inflammation is restricted to the mucosa and submucosa (the two innermost layers) of the intestinal wall and only affects the large intestine. Surgery is needed in up to 15% of patients at 10 years.¹⁹ Between five and 10% of patients with a diagnosis of UC have their diagnosis changed to CD.

The disease course in UC and CD is affected by age at diagnosis, extent and location of disease and lack of endoscopic healing while in remission.^{19, 20}

IBD patients are at a greater risk of developing certain cancers. For instance, UC patients have a 2.4-fold higher risk of developing colorectal cancer (CRC) and CD patients have a 67-fold increased risk of developing small bowel cancer compared to the general population.²⁷

Co-morbidities for UC and CD patients include arthritis, asthma, bronchitis, psoriasis, and pericarditis.²⁸ Mortality is 50% greater among CD patients than among the general population but patients with UC do not have an overall increased mortality.²⁷

1.2.2.2.3 Aetiology and pathophysiology - underlying causes and effects

The aetiology of CD and UC is complex and multifactorial most likely leading to an inappropriate immune response to the normal and or altered intestinal microbial flora.²⁹

About 12% of patients have a family history of CD.²⁰ Genetic factors play a smaller role in UC (7.5%).²⁹ Other important factors include epigenetic and environmental factors such as smoking, diet, micronutrients and some medications.²⁰ For many environmental factors the causative association remains to be proven. Furthermore, a change in microbiota is thought to be a factor in the cause of CD provoking the secretion of high amounts of tumour necrosis factor alpha, a cell signalling protein (cytokine) involved in systemic inflammation.

1.2.2.2.4 Diagnosis

The diagnosis of IBD relies on symptom assessment by gastroenterologists, endoscopy with histological findings and the exclusion of alternative diagnoses (for instance infectious colitis, intestinal tuberculosis, Behcet's disease or coeliac disease).^{19, 20} Inflammatory markers, such as C-reactive protein and faecal calprotectin, may be useful indicators for endoscopy but are in their own right not helpful in the differential diagnosis of UC and CD. Therefore, a definitive diagnosis requires colonoscopy with biopsies taken from at least 6 sites.¹⁹ In addition small bowel capsule endoscopy can identify small bowel involvement in CD when endoscopy studies have been negative. Computed tomography or magnetic resonance imaging can be useful at the time of diagnosis to assess the extent of disease and presence of complications.

The median time (and interquartile range [IQR]) from first presentation of symptoms to definitive diagnosis is nine months (IQR 3-24) for CD and four months (IQR 1-12) for UC.³⁰ Delayed diagnosis of IBD reduces treatment options and increases the risk of disease progression.³¹

Following diagnosis, disease needs to be sub-classified and its activity assessed to define the extent and severity of IBD.^{19, 20, 32} This is important for selecting the most appropriate treatment.

1.2.2.2.5 Management

Specialist treatment, mainly in the form of drugs, is usually initiated by a gastroenterologist following referral from general practice. Drug treatment of IBD involves an induction (induce remission) and maintenance (maintain remission) regimen. The choice of treatment depends on disease severity and response to previous therapies. The most widely used drugs in IBD are 5-aminosalicylic acid, corticosteroids, immunomodulators (thiopurines and methotrexate), and biologicals (such as infliximab, adalimumab, vedolizumab). In addition smoking cessation programmes and nutritional therapy form part of the management of CD while probiotics and faecal transplantation have no established role yet.²⁰

The treatment target for IBD has evolved from simple symptom control to deep remission (mucosal healing on endoscopy).^{3, 33} Deep remission is the treatment target of choice because it is associated with reduced relapse rates, reduced need for surgery, and less bowel damage.²⁰ Colonoscopy plays an important role in

monitoring disease activity and treatment response to achieve the treatment target. However, faecal calprotectin is increasingly being used to monitor disease activity because concentrations correlate well with endoscopic activity while symptoms do not. Therefore, symptoms alone should not generally guide therapeutic decisions.²⁰

Once treatment is established, three different models of care of IBD patients exist: 1) hospital based care ideally in a specialised IBD clinic, 2) shared care systems with primary care, and 3) patient managed care.³⁴ Effective communication between primary and secondary care is essential for the optimal management of patients with IBD especially as GPs are becoming increasingly involved in the shared care of patients. However, in a 2013 UK survey of IBD patients aged 29 years or younger it was the secondary care IBD nurse specialist who had an increasing role as the first point of contact during a flare and in the communication about management.³⁵ The IBD nurse specialist forms a key part of the integrated multidisciplinary team in the management of IBD patients and in 2013 76% of IBD patients had access to a nurse specialist.³⁵

In patients who develop complications or in whom treatment fails surgery is considered.

1.2.2.3 Summarising IBD and IBS in comparison

IBD and IBS are chronic diseases of the intestinal tract and common symptoms include chronic diarrhoea and abdominal pain. IBS is a non-progressive, non-organic disease, usually managed with dietary and lifestyle advice in primary care. IBD is an organic disease caused by inflammation of the intestine. The disease is severe and progressive. Patients require drug treatment following confirmative diagnosis in secondary care using colonoscopy. The majority of patients require surgery at some stage. The difference in severity and management of IBD and IBS requires differential diagnosis, and it is important not to miss patients with IBD because a delay in treatment adversely affects outcomes. However, given the considerable overlap between symptoms, some have argued that IBS and IBD are clinical presentations on a spectrum of disease whereby IBS symptoms are possibly early signs of IBD.³⁶ Supporting this view is the 9-16 times higher rate of IBD in IBS patients than in the general population.¹² The average time between IBS and IBD diagnosis is 2–3 years.^{12, 36} Another possible explanation may be that the symptom overlap leads to misdiagnosis of IBS in patients with IBD or microscopic colitis. Furthermore, prevalence of IBS-type symptoms in IBD patients without ongoing

intestinal inflammation is between 8% and 28% suggesting that the two conditions can co-exist further hindering the differential diagnosis.³⁷ There are close to 300,000 patients with IBD in the UK²³ compared to an estimated four million people with IBS.⁶ General practitioners refer between 10-20% of patients presenting with abdominal symptoms if no specific test is available.^{38, 39} However, only about 25% of referred patients have organic disease, of which a third have IBD.^{40, 41}

1.2.3 TESTING PATIENTS WITH ABDOMINAL SYMPTOMS

1.2.3.1 Diagnostic strategies to differentiate IBD from IBS

The overlap of symptoms between IBD and IBS, but also more broadly between organic (including IBD) and non-organic (mainly IBS) intestinal conditions, results in diagnostic uncertainty and unnecessary referral.^{17, 42, 43} This is compounded by the lack of a diagnostic test for IBS.¹⁵ Diagnosis of IBS by excluding organic conditions through extensive testing is costly and can produce false positive (FP) results (incorrectly classify IBS as an organic condition). Therefore, a single, reliable and non-invasive test would have considerable clinical value.¹⁵ However, available clinical criteria and indices are subjective and lack accuracy.⁴⁴ Serological markers including levels of C-reactive protein and erythrocyte sedimentation rate are non-specific to bowel inflammation, and endoscopy and other imaging technologies are often invasive, expensive and involve long waiting times. Faecal inflammatory markers that either leak from or are produced by the gastrointestinal mucosa during inflammation are a better option.⁴⁴ A number of candidates have been studied, of which lactoferrin and calprotectin showed the most encouraging diagnostic accuracies.⁴⁴ Calprotectin is the most widely studied.⁴⁴ Low levels of faecal calprotectin can aid the exclusion of intestinal inflammation but not specific organic conditions because not all conditions are always characterised by mucosal inflammation.

1.2.3.2 Faecal calprotectin

Faecal calprotectin is a small calcium-binding protein of the S100 family.⁴⁴ It is predominantly derived from neutrophils, which are the inflammatory cells first to migrate towards the site of an inflammation following chemical signals. FC is a major protein in the cytosol of neutrophils making up 60% of the soluble proteins.⁴⁴ Calprotectin plays a regulatory role in the inflammatory process and stops bacteria and fungi from reproducing.⁴⁵ In the presence of calcium it is stable up to seven days at room temperature and it is resistant to enzymatic digestion.⁴⁶ The release of calprotectin is proportionate to the degree of inflammation which makes it an ideal

marker for the level of inflammation in inflammatory diseases. In 1992, Roseth et al.⁴⁷ developed the first biochemical test for isolating and quantifying FC. Significant correlations of FC and intestinal inflammation could be shown for ulcerative colitis and Crohn's disease.^{48, 49} Subsequent studies reported significantly higher FC levels in patients diagnosed with IBD (median 1251µg/g, IQR 532-2325µg/g) than those with a functional diagnosis (median 20µg/g, IQR <20-50µg/g).⁵⁰⁻⁵² Levels in patients with other gastrointestinal conditions were intermediate (median FC 50µg/g, IQR 20-145µg/g).⁵⁰ Other conditions that cause raised levels of FC are coeliac disease, infectious colitis, necrotizing enterocolitis, intestinal cystic fibrosis and CRC.⁵³

Banerjee et al. (2015) investigated different FC thresholds to exclude inflammation in secondary care and suggested 50µg/g as the optimal threshold when the clinical objective is to distinguish between IBD and IBS.⁵⁴ They achieved sensitivity of 100% and spared 52% of referred patients having colonoscopy. Lower thresholds produced poorer specificities.⁵⁴ For primary care, a higher threshold of 100µg/g has been proposed with the aim to improve the positive predictive value.⁵⁵

1.2.3.3 Evaluation of FC as a diagnostic test

Test evaluation requires the assessment of test accuracy, the impact of testing on clinical decision making and the clinical effectiveness and cost-effectiveness of patient management following testing.^{56, 57} Clinical utility of a diagnostic test is only achieved if testing improves diagnosis, management and outcome of patients.⁵⁷ Test accuracy assesses the test's diagnostic performance (how often the test is correct or incorrect). An increase in test accuracy cannot be assumed to improve patient outcomes or cost-effectiveness as the effect of test accuracy on clinical decision making also needs to be evaluated.⁵⁶

Tests are either evaluated in (rare) test-treatment trials or more commonly in decision analytic models which first link test accuracy evidence with clinical effectiveness data (linked evidence approach) and then extrapolate long term outcomes under a number of assumptions.⁵⁷ The evaluation of the FC test used the linked evidence approach.⁵⁸ A systematic review of FC test accuracy studies estimated the sensitivity and specificity of FC testing to distinguish IBD from IBS to be 0.93 (95% CI 0.83 to 0.97) and 0.94 (95% CI 0.73 to 0.99), respectively.⁵⁸ This was linked with evidence from the literature on the clinical effectiveness of IBD management. The subsequent economic model required a number of critical assumptions on test use and test impact on clinical management.⁵⁸ Based on the

evaluation by Waugh et al. (2013)⁵⁸ NICE recommended FC testing for primary care in 2013 to “support clinicians with the differential diagnosis of IBD or IBS in adults with recent onset lower gastrointestinal symptoms for whom specialist assessment is being considered” if colorectal cancer is not suspected.⁵⁹

1.2.3.4 Summarising the diagnostic potential of faecal calprotectin

Faecal calprotectin is a marker of gastrointestinal inflammation. Levels are generally raised in patients with IBD but not IBS. As the protein is not digested in the intestine it can be quantified in stool samples presenting a relatively simple test for inflammation (but not specifically for IBD). It is more specific than serum inflammatory markers. FC testing is established in secondary care to diagnose and monitor IBD patients. In 2013 FC testing was approved by NICE for the differential diagnosis of IBD and IBS in adult patients with abdominal complaints in primary care.⁶ In clinical practice this would require the exclusion of other conditions such as red flags for suspected CRC and coeliac disease before testing a population enriched in IBS and IBD cases. The decision was based on an evaluation of FC testing using the linked evidence approach and economic model under a number of assumptions. These assumptions formed the rationale for my investigations.

1.2.4 LACK OF EVIDENCE FROM PRIMARY CARE – THE RATIONALE FOR THIS THESIS

The approval of FC testing in primary care was motivated by the expectation of reducing National Health Service (NHS) costs for unnecessary referrals to secondary care by over 60%.⁶ At the time no test accuracy studies from primary care were available; further, it was not known how the FC test would impact on clinical decision making. Thus, the positive cost-effectiveness outcome in the diagnostic assessment report relied on two major assumptions.

1.2.4.1 Assumption 1 (patient management): referral of only FC positive patients

The evaluation assumed that GPs would only test those they would otherwise have referred without testing, and only refer patients with a positive FC test to secondary care.⁵⁸ The evaluation concluded that FC testing results in cost savings because of a reduction in colonoscopies.⁵⁸ However, subsequent evidence suggests that cost savings may be overly optimistic.

There is consistent evidence that around 70% of FC tests in primary care are normal (negative results).^{51, 60-64} However, between 26% and 66% of these patients

were referred to secondary care by the primary care clinician, and between 40% and 70% of these referred patients received colonoscopy.^{51, 60, 63-65} The extent of test negative referrals and the impact this may have on cost-effectiveness claims and patient care in terms of opportunity cost is not known. Consequently, it is uncertain whether the introduction of FC testing into primary care can translate into the NHS use and patient benefit originally envisaged.⁶

1.2.4.2 Assumption 2 (test accuracy): sensitivity and specificity are the same in primary and secondary care settings

The recommendation for use of the FC test in primary care⁵⁹ was based on test accuracy studies in secondary care, assuming that test characteristics remain constant between primary and secondary care settings. Irwig et al. (2002) list a number of requirements to be met before diagnostic test accuracy estimates can be transferred between populations.⁶⁶ These include, that the definition of disease, the test used, the thresholds, and the prevalence are all constant. Several of these do not hold in the case of FC testing. Most importantly, the patient population is different between the two settings with a different patient spectrum and IBD prevalence in the primary care population.⁶⁷ While test sensitivity and specificity are mathematically independent of disease prevalence (proportion of IBD cases), the spectrum (the range of severity of disease symptoms) among tested patients impacts on test accuracy estimates.⁶⁸ The disease spectrum is linked to prevalence. For instance, there are many more patients with IBS than with IBD in primary care resulting in a wide range of FC values. Furthermore, it is likely that a greater proportion of mild IBD cases are present in primary care with FC levels closer to normal. The secondary care population resembles only a subpopulation of the primary care population and the disease does not span the whole spectrum of severity found in primary care as a result of the filtering by the GPs. These are potential reasons why test accuracy estimates may vary in primary care and transferability of test accuracy data cannot be assumed. The direction or magnitude any difference in test accuracy might take and what impact this may have on the cost-effectiveness of FC testing in primary care requires investigation.

1.2.4.3 Summarising the rationale for my thesis

The FC test was approved by NICE in 2013 for use in primary care following a thorough high quality systematic review and health economic evaluation.⁵⁸ However, the decision was solely based on evidence of test accuracy from secondary care. Transferability of test accuracy estimates between settings cannot

be assumed when patient populations differ between settings.⁶⁸ Assumptions of appropriateness of testing and referral in the diagnostic assessment report were overoptimistic. There was some evidence that between 26-66% of patients with a low FC level may be inappropriately referred of which 40-70% received colonoscopy.^{51, 60, 63-65} Understanding the role of FC testing in clinical practice is a prerequisite for a reliable cost-effectiveness analysis. In my PhD I, therefore, investigated the role of FC testing in primary care to address the following aim and objectives.

1.3 Aim and Objectives

The aim of this research was to establish primary care evidence on the test use and test accuracy of FC testing in order to discuss the validity of current health economic estimates for the primary care context. To achieve the research aim, I addressed the following main objectives:

- 1) To characterise the uptake and use of FC testing in primary care and investigate the referral of tested patients,
- 2) To estimate the test accuracy of FC testing in primary care and
- 3) To explore influences on GPs' decision to refer in order to explain observed referral patterns following FC testing.

I devised five studies to address these objectives. The thesis outline explains how these studies are situated in my thesis.

1.4 Thesis outline

In this first chapter, I provided an overview of the two conditions of interest (irritable bowel syndrome and inflammatory bowel disease), the FC test, its evaluation and the rationale for my investigations. I explained that missing primary care evidence for the evaluation of FC testing prompted me to investigate the current use and test accuracy of FC testing in primary care.

In Chapter 2, I discuss my chosen methodologies, a systematic review, a retrospective cohort study and qualitative interviews, in the context of my wider approach; that is the research paradigm with its ontological and epistemological underpinnings. I explain what this means in terms of the interpretation of study

findings and the question of study validity. I present a justification for the suitability of the methodologies to address my research questions.

In Chapter 3, I investigate the use of FC testing in current clinical practice. I consider the advantages and disadvantages of routine data for research. I describe the Health Improvement Network (THIN) database (now known as IQVIA Medical Research Database (IMRD-UK)), which is a database of electronic patient records made up of information routinely collected as part of GP consultations. This is followed by a descriptive study using the THIN database to establish a) the incidence and prevalence of IBD in routine primary care electronic health records of adult patients, b) the uptake of FC testing in primary clinical practice, and c) the extent of referral of test positive and negative patients nationally. The main aim of the study was to understand and characterise FC test use in primary care since it was recommended by NICE in 2013. In particular, I aimed to estimate the extent of FC test negative referrals, which could potentially reverse the positive cost-effectiveness outcome.

In Chapter 4 I focus on test accuracy. First, I describe my systematic review of test accuracy studies of faecal calprotectin to establish summary estimates of sensitivity and specificity for primary and secondary care. Then, I report my test accuracy study using the GP database. The study included patients with at least one FC test result and no prior IBD diagnosis. FC test results were verified with information provided in the database. I estimated sensitivity and specificity estimates that were applicable to the primary care setting. Finally, I describe my tailored meta-analysis. This method allowed me to identify those studies for meta-analysis which report sensitivity and specificity estimates which are plausible for the primary care setting. I did this by assessing published studies against information on the IBD prevalence and FC test positive rate obtained from the THIN database. I compare the results with estimates from the systematic review and test accuracy study.

In Chapter 5, I report my qualitative interview study with 19 GPs, in which I explore GPs' views and experiences regarding referral and what influences their referral decisions. The main aim was to understand who GPs refer and why, and where the FC test fits into the referral pathway. I use the findings to explain the apparently inappropriate referral decisions of FC negative patients, which the economic evaluation of FC testing failed to predict.

In Chapter 6, I revisit the existing health economic model and discuss the model structure, its assumptions, base case parameters and sensitivity analyses in light of my outcomes from Chapters 3 and 4. I frame my claims based on the insights gained in Chapter 5 to allow me to assess the applicability of cost-effectiveness results to current clinical practice in primary care.

In Chapter 7, I summarise my contribution to research and reflect on the study findings in light of the three main study objectives. I bring together findings from all my studies to discuss the test accuracy and use of FC testing in primary care considering the NICE guidance. I further discuss implications for clinical practice and future research needs.

Chapter 8 is my conclusion. I present my learning from the study findings and my opinion on the usefulness and cost-effectiveness of FC testing in primary care.

I use the visual thesis outline in Figure 1.1 to highlight the position of the chapters in the thesis.

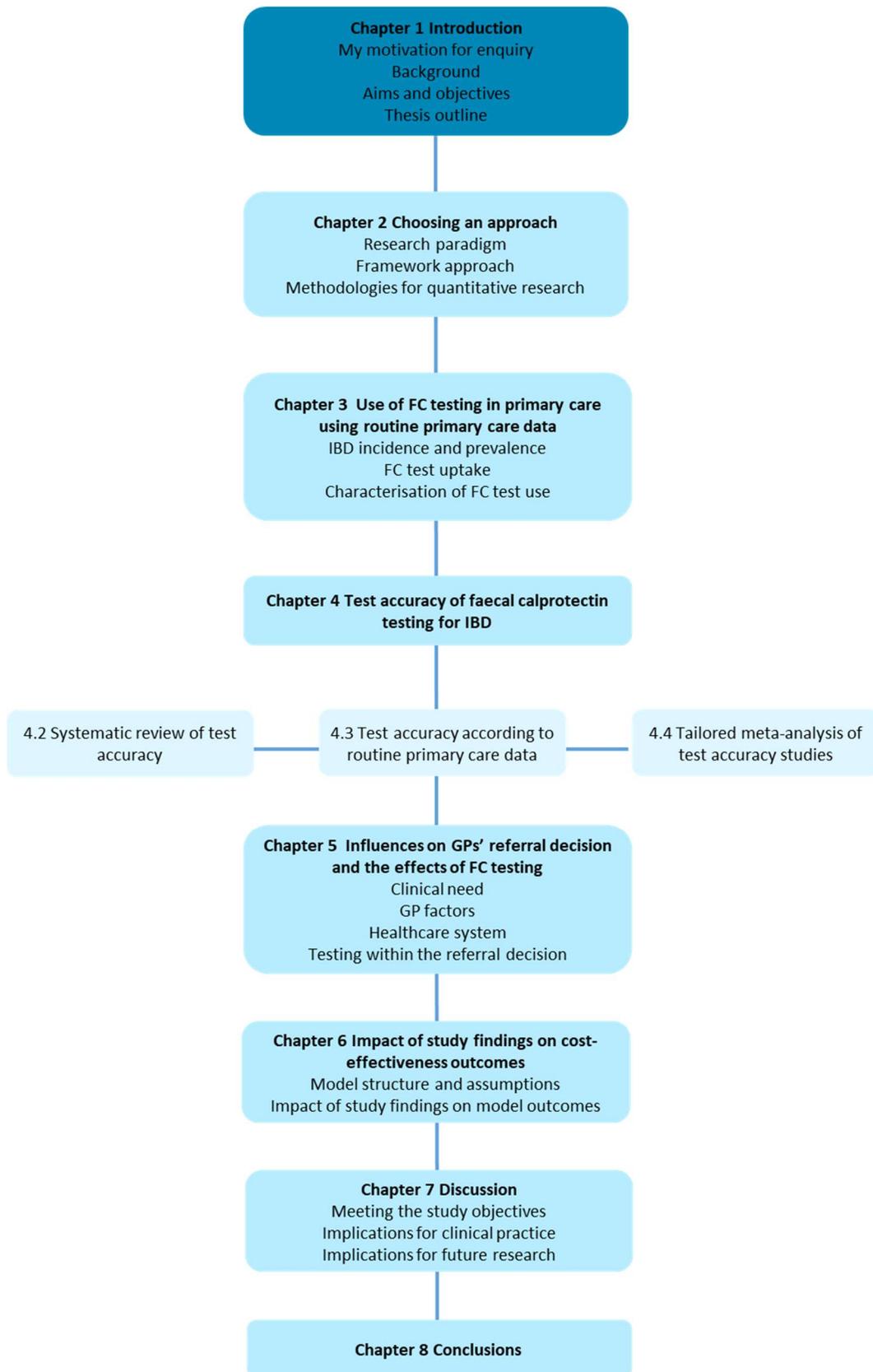
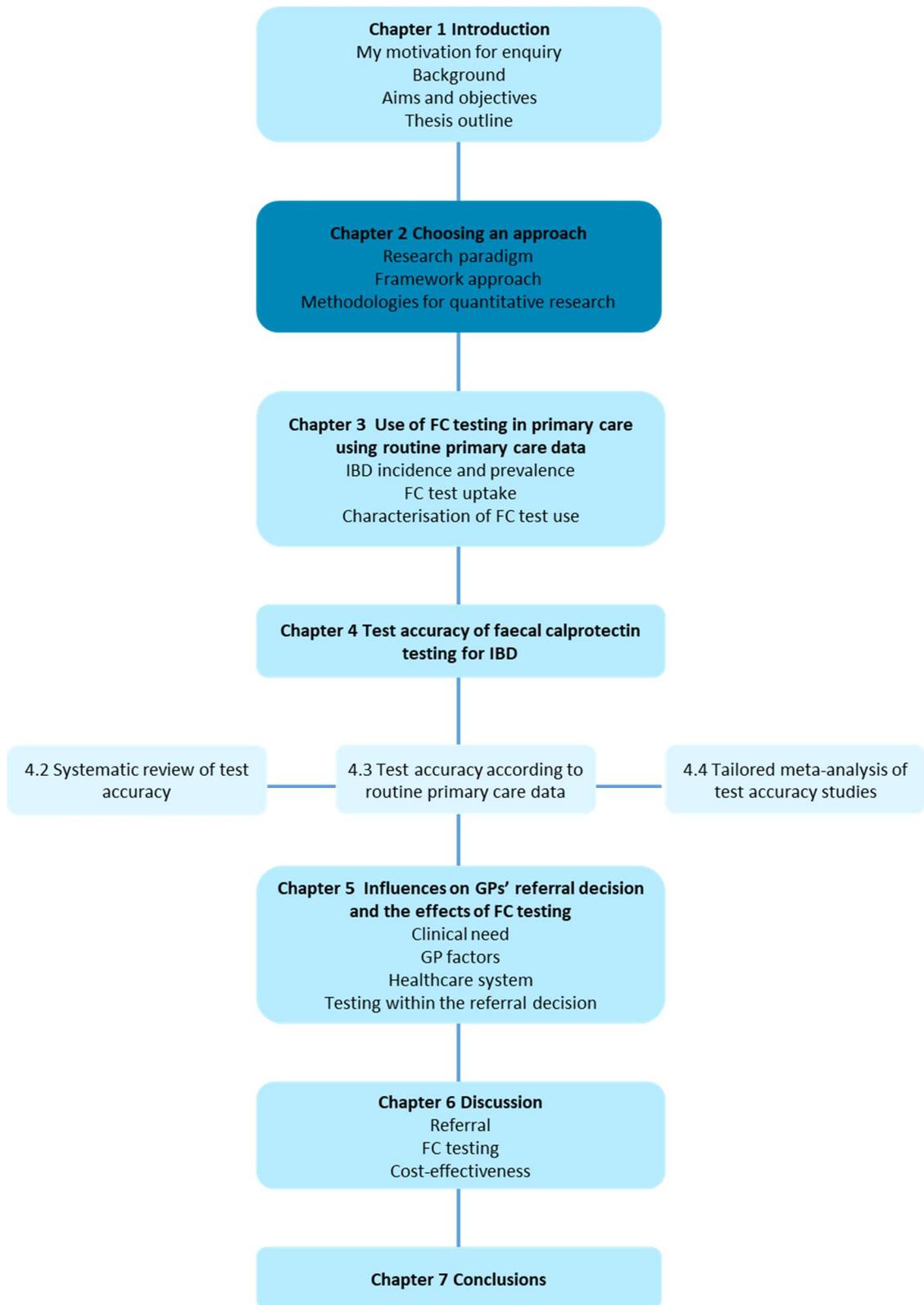


Figure 1.1 Visual outline of thesis structure



Chapter 2 Choosing an approach

2.1 Chapter overview

The overall aim of this thesis was to discuss the validity of health economic estimates for FC testing in primary care in light of new evidence on test use and test accuracy. I chose my methodologies considering this study aim. I collected quantitative and qualitative data to gather a broader perspective on the research issue. I used a quantitative approach to investigate the uptake of testing, test accuracy and referral decisions following testing. I used qualitative data collection to aid the explanation and understanding of findings in the context of clinical practice. I used the different methodological approaches to address separate research questions. Therefore, I did not formally mix methods or results. Instead, I combined findings in the discussion of the cost-effectiveness results in Chapter 6.

Methodological approaches in research are underpinned by philosophical assumptions, which determine the degree to which a research study is accepted and by whom.⁶⁹ This chapter aims to set the scene for my research by describing the research paradigm that I subscribe to, based on my research questions. By doing so, I take the view that the research should guide the methodology⁶⁹⁻⁷¹ and that trying to fit the research to certain theoretical traditions may obstruct choosing the most appropriate research design to answer the research questions posed.⁶⁹ I explain my interpretation of where this research sits in the arena of ontological and epistemological debate and provide a justification for my three chosen methodologies; framework analysis,⁷² systematic review and the retrospective cohort study.

2.2 Research paradigm or world view

In my research I investigate the potential shortcomings identified in the evidence base of a national clinical guidance. The research has an applied focus and is not entirely theoretical. I take the view, for instance, that it is important to understand referral behaviour in order to be able to influence it. Referral also has very real consequences for patients and the NHS. I, therefore, take a realist perspective in this thesis.

Guba and Lincoln (1994)⁷³ argued that the epistemological stance - how we know what we know - is constrained by ontological beliefs - what there is to know about

the world - and vice versa. That is, if the researcher believes that there is an objective truth to be known (realist ontology) the researcher is taking an objective epistemological stance to discover value free how things really are.⁷³ However, research requires many subjective decisions and my interview findings clearly depend on my personal interaction with the GPs. My epistemological question is, therefore, how close can I come to the truth by talking to 19 GPs and using routine data to address the research questions? I would argue that I can only strive to get to the truth by choosing robust research methods. These include 1) talking to a range of GPs using maximum variation sampling; 2) practice reflexivity throughout the research process; 3) being transparent about study assumptions; and 4) considering quantitative as well as qualitative data to achieve a wider range of answers and a broader perspective on the overall issue. Qualitative and quantitative data have equal status in my research. With that I take the stance that qualitative and quantitative data are compatible.^{70, 74} The combination of a subjective epistemology with a realist ontology and the compatibility of qualitative and quantitative research goes against the purists' views including Guba and Lincoln.⁷³ However, others argue that there is a need for some middle ground.^{74, 75} My approach of combining qualitative and quantitative data is well placed within the broad philosophical school of thought generally known as 'critical realism'.⁷⁵ By subscribing to the critical realist paradigm I see reality as something that exists independent of me studying it (ontological realism) and my research findings are created by my interaction with the 'object of investigation' (epistemological constructivism).⁷³ However, my ability to know this reality is imperfect and 'objectivity' is an ideal.⁷⁶ Choosing rigorous methodological strategies can help in closing in on that ideal.⁷⁶ Within the critical realist paradigm I am not committed to particular methodological traditions^{69, 74} and I chose my methodologies considering the aims and objectives of my studies.

2.3 Framework approach for qualitative data

The framework approach fitted the aim of my research to provide actionable outcomes and recommendations rather than to generate social theory.⁷² It offered a systematic and transparent approach for my analysis of semi-structured interviews,⁷⁷ and the systematic method in organising and summarising data allowed me to share and discuss findings at different stages of analysis with patient as well as academic advisors. Matrices developed during data management provided a clear audit trail of how results were obtained from the data and enabled a

simple way of comparing and contrasting data across and within individual cases.⁷⁸ The methodology is not aligned with a particular philosophical paradigm but is a flexible tool that enables for instance the combination of inductive and deductive approaches to inquiry.⁷⁸

2.3.1 INDUCTIVE-DEDUCTIVE BALANCE

In my interviews I took an inductive approach. I asked GPs open questions that were not designed to test a hypothesis but to encourage GPs' perspectives and stories to emerge. For the same reason, coding started with open coding rather than using codes derived from existing literature. Even though I believe that knowledge is best acquired through this bottom up process, I need to acknowledge the fact that there were deductive aspects in my learning throughout the project. Firstly, the rationale for the research was based on the observation of inappropriate referral decisions. Secondly, the research question was refined through reading of relevant literature. Finally, I compared research outcomes to other findings in the literature. My research has, therefore, a deductive as well as an inductive aspect. By choosing the framework approach,⁷² I could turn this to my advantage and combine previous knowledge with emerging knowledge in the stages of analysis.

2.3.2 VALIDITY AND GENERALISABILITY

There has been much debate whether qualitative research can and should be validated in a similar way to quantitative research findings using validity, reliability and generalisability as measurements of quality.⁷⁹ One of the issues is that it assumes the existence of a true reality against which the findings can be measured.⁷⁹ Within the critical realist paradigm, which assumes that there is an objective reality, the concepts of reliability and validity are relevant.⁸⁰ However, our knowledge of this reality is imperfect. Angen (2000)⁷⁹ suggested that the notion of validity as truth or certainty must be abandoned and that research should aim for trustworthiness. Strategies including member checking, peer review, external audits and reflexivity have been suggested for producing more credible and rigorous research.⁷⁶ The systematic method of organising and summarising data in the framework approach⁷² lends itself particularly well for these strategies.

Using the framework approach⁷² also means sharing the belief that qualitative research is generalisable in terms of the range and diversity of the findings but not in relation to their prevalence.⁶⁹

2.4 Methodologies for quantitative data

2.4.1 SYSTEMATIC REVIEW

Systematic reviews are the cornerstone of evidence-based practice in order to identify, evaluate and summarise the findings of all available research evidence and the best way to begin any new primary study.⁸¹ It was, therefore, my first choice of methodology in the evaluation of test accuracy before undertaking a new study. A systematic review provides an objective appraisal of the evidence, while meta-analyses enhance the precision of summary estimates.⁸² The aim of my review was to explore the differences in test accuracy estimates between primary and secondary care studies and to identify studies for tailored meta-analysis.

2.4.2 RETROSPECTIVE COHORT STUDY

Test-treatment trials are the highest level of evidence in the hierarchy of study designs for test accuracy.⁸³ I did not consider a test-treatment trial for three reasons. Firstly, IBD is a rare condition and test-treatment trials are often too small to detect patient health effects.⁸³ Secondly, clinical equipoise did not exist to allow a controlled design because the test was already approved by NICE. Thirdly, findings from controlled trials may have limited generalisability to general practice.⁸¹ My aim was to produce evidence that is applicable to clinical practice and can support decision making. This can be achieved with a pragmatic study design.⁸⁴ I, therefore, chose the cohort design to assess the performance of FC testing in a large, multicentre study population representative of the setting and clinical question of interest. I recognise that test accuracy studies are cross-sectional in nature, but I use the term 'cohort design' because patients are followed-up over time to establish their final disease status. Furthermore, this terminology makes it simpler to distinguish between prospective and retrospective designs.

A prospective cohort would have been favourable for the collection of primary care data for the assessment of FC test accuracy. This would have allowed me to collect data on patients' symptoms, reasons for testing, FC test results and follow-up data on additional investigations and final diagnosis for each patient. However, such study was not feasible for my PhD. IBD is a rare condition and a single GP practice would provide too few new cases for analysis. Widening the study to several GP practices would have required a mobile research nurse for consenting and data collection. Furthermore, such a study would not be justifiable. The reference standard to verify FC test outcomes is colonoscopy and it would be unethical to

subject patients to colonoscopy where disease is not suspected. An appropriate alternative reference standard could be generated from follow-up of test negatives. Such data are readily available in routine GP databases. Routine electronic patient records are particularly useful for studying what is happening in the real world and offer speed of access to population size data without the need for expensive data collection.⁸⁵ I, therefore, chose to undertake a retrospective cohort study of routinely collected primary care data to produce estimates of sensitivity and specificity of FC testing in a primary care population and to study the uptake and use of FC testing. I recognise this study design has limitations for my research including reliance on coded information and missing data.

2.4.3 VALIDITY AND GENERALISABILITY

The systematic methods of the systematic review render the review process transparent and reproducible. Quality appraisal assesses the bias of studies included into the review and their applicability to the research question. This supports an honest discussion of the validity and generalisability of the overall review findings.

The analysis of routine data required a number of subjective assumptions for the definition of variables and follow-up times and for the inclusion of patients into the study analysis. This may compromise the validity and reliability of study findings. However, I used a transparent method of reporting assumptions and study limitations which supports readers in the judgement of the validity of the study findings. The use of routine data aids the generalisability of study findings. By using descriptive and inferential statistics I aimed to generalise, infer and predict from the study sample to the population from which it was drawn (external generalisation).⁷⁴

However, within the critical realist approach I recognise that my knowledge from the studies is fallible and incomplete.⁷⁵ This uncertainty in the quantitative study findings is reflected by reporting p values and confidence intervals (CI).

2.5 Summarising my overall research approach

I decided on a research approach based on my study aim and objectives. My individual studies and my overall research aim sit well within the critical realist paradigm which supports the combination of qualitative and quantitative research methods to gain a more complete understanding of the research question at hand

and a broader perspective on the overall issue. My overall approach as depicted in Figure 2.1, makes use of two data types: quantitative data from systematic review and meta-analysis of published studies (Chapter 4) and routine electronic health records (Chapter 3 and Chapter 4), and qualitative data from semi-structured interviews (Chapter 5). I use the different data types to address separate research questions which prevents formal mixing of methods or results. However, I combine findings in the discussion of the cost-effectiveness results in Chapter 6 and in the overall discussion in Chapter 7 to enhance and clarify findings from the quantitative analysis with those from the qualitative analysis.

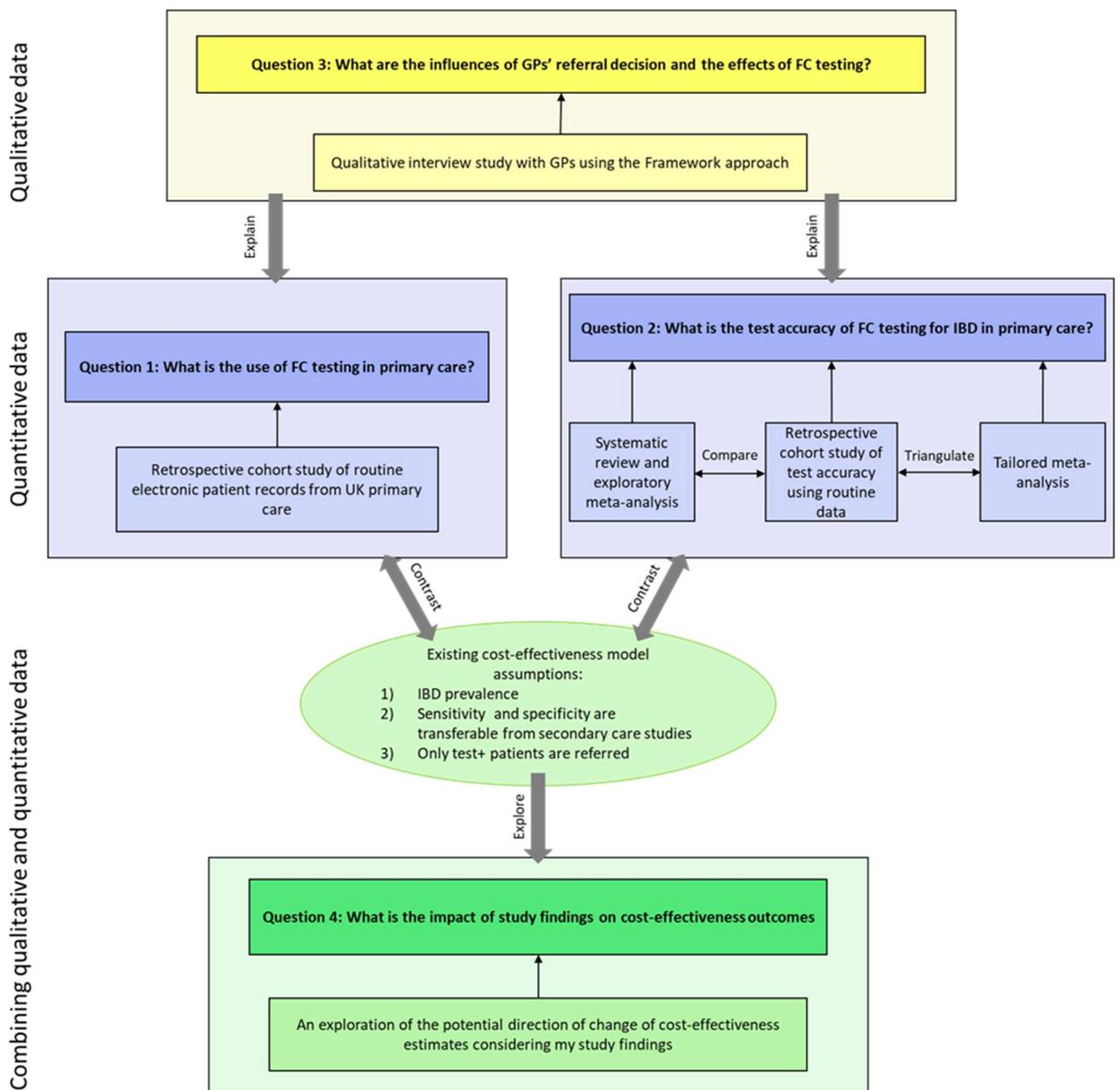
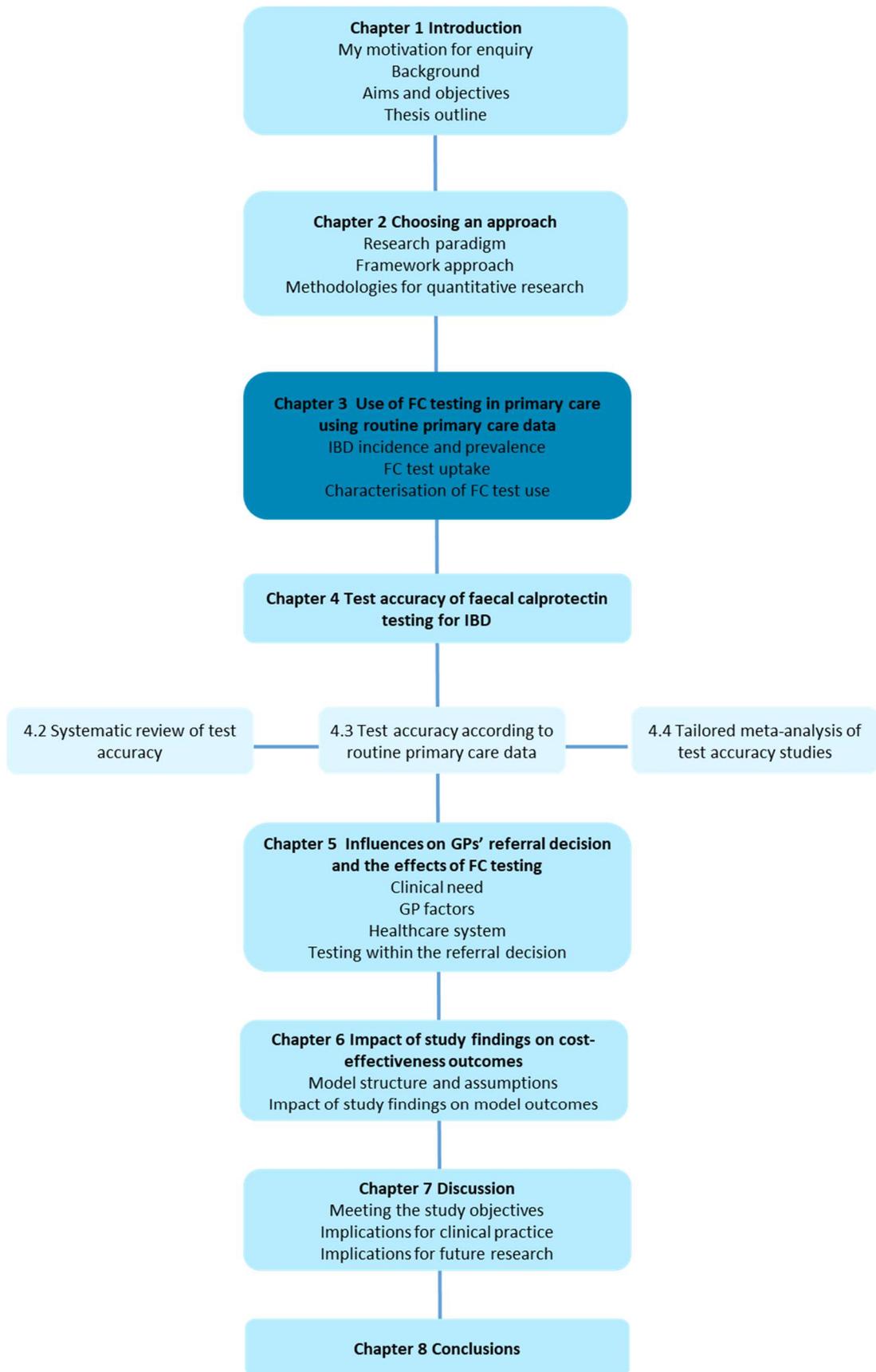


Figure 2.1 Logic model depicting my approach of investigation



Chapter 3 Use of FC testing in primary care

3.1 Chapter overview

This chapter provides an introduction to using routine data for research more generally while in section 4.3.1 I talk about routine data for test accuracy studies. The main aims of this chapter are 1) to describe the THIN database, a database of primary care routine electronic health care records, 2) to explain the methods used to create THIN datasets for my specific research questions and 3) to explore the data on the use of FC testing in primary care. My research on the use of FC testing in primary care covers four areas:

- 1) the incidence and prevalence of IBD in the THIN population,
- 2) the uptake of FC testing into primary care clinical practice,
- 3) the characteristics of FC tested patients and
- 4) the effect of FC testing on patient management in terms of referral and colonoscopy following FC testing.

First, however, I review the literature on using routine data for primary care research and discuss the appropriateness and usefulness of using the THIN database to address these four research areas.

3.2 Introduction

3.2.1 ROUTINE DATA FOR PRIMARY CARE RESEARCH

3.2.1.1 UK primary care databases

UK primary care data are unique and particularly suitable for research for the following reasons: 1) Over 95% of the UK population is registered with a GP^{23, 86} and access to health care is free of charge at the point of delivery. This results in comprehensive coverage of the population included in the database. 2) GPs act as gatekeepers to all services and specialists in secondary care (excluding emergency care). This means that a referral from a primary care professional is required in order to access secondary or tertiary care and this is recorded in the patient's record. 3) Patients are usually only registered with one GP at any one point in time; and for each patient the registration date and the date when the patient leaves the practice is known. This provides longitudinal data with known start and end date of follow-up. 4) The role of the GP extends to the management of chronic patients. This means that the database includes information on tests and lifestyle behaviour.

In the mid-1980s UK primary care started the transition from paper to electronic health care records producing a centralised health record for all patients.⁸⁷ At the same time providers of the clinical computer systems established electronic health records databases and secondary use of the collected data for research started.⁸⁸

There are four UK primary care databases that are well established in the global research community and have contributed to a substantial number of scientific studies and publications.⁸⁹ These are the Clinical Practice Research Datalink (CPRD), THIN, QResearch and ResearchOne, which collect data from GP practices using different clinical computer systems.⁸⁸⁻⁹⁰ ResearchOne is the smallest and QResearch the largest database.⁸⁸ Access to QResearch (which moved from the University of Nottingham to the University of Oxford in 2019) is restricted to on-site research and a maximum of 100,000 patients. There is considerable overlap between practices in the CPRD and THIN with higher service costs associated with the CPRD database.⁸⁸ I chose the THIN database because my established collaboration with the University of Birmingham meant access to the database through their annual sublicense and to their research support. THIN has collected patient data since 2003 from practices subscribing to the Vision software.⁸⁸

3.2.1.3 Advantages and limitations of routine primary care databases for research

Routine data are cheaper and more instantaneous than prospective data collection.⁸⁸ The size of the database means that rare conditions, such as IBD, can be studied more readily.⁹¹ As the THIN population is a defined population all patients with a given disease can be investigated. Further strengths include the availability of prescription and laboratory data.

General limitations of routine electronic health records databases are linked to the fact that the data are collected with the management of patients in mind rather than primarily for research. This may result in missing data, which I will consider in greater depth in the discussion alongside the study findings (section 3.6.2.3). Limitations may also be linked to the geographical representativeness of the data, which is determined by the location of practices using the clinical computer system from which the data are drawn.⁸⁸

In 2016, only 9% of UK GP practices used the Vision computer software which is the clinical computer system for the THIN database.⁸⁸ And even though the practices covered 45.9% of the Clinical Commissioning Groups (CCGs), they were

clustered mainly in London, the South, Greater Manchester and Birmingham with limited coverage of the North and East of England.⁸⁸ However, the actual geographical coverage by the THIN data might differ slightly from the geographical representation of the Vision software because not all practices contribute data. I do not know the reasons why practices decide to contribute / not contribute to THIN. I, therefore, consider studies, which validated the THIN database and showed the representativeness of the THIN population to the UK population.

Studies using the THIN database report that the THIN population is broadly representative of the UK population and findings can be generalised to the broader UK population. Blak et al. (2011) reported that chronic disease prevalence in THIN is comparable to national rates albeit generally slightly higher. They explained this finding by using crude condition estimates for comparison not controlled for age and gender and by differences in practice software and GPs' recording practices.⁹² They also reported that the THIN population and the overall UK population were similarly distributed across age and gender, although, THIN contains slightly fewer people aged under 25 years.⁹²

The completeness of data recorded by Vision practice staff was confirmed by assessing and comparing consultation, death and prescribing rates to rates from the General Household Survey, the Office for National Statistics and rates reported by the Department of Health, respectively.⁹³ Furthermore, the accuracy of recorded data was confirmed by looking for female specific conditions in males and age specific conditions, such as dementia, in children.⁹³ Data recorded with the Vision software were rated as of sufficient quality for use in medical research.⁹³

3.2.1.4 How is THIN organised?

The THIN database consists of longitudinal individual level patient data on age, sex, medical diagnosis, symptoms, health promotion activities, referrals and prescriptions. In 2015, THIN collected data from more than 670 GP practices and a total of over 14 million patients had contributed data to THIN.⁸⁸ This reflects a coverage of about 6% of the UK population.⁸⁸

When a practice joins THIN, the complete set of retrospective patient records is sent to THIN, followed by daily, automatic downloads of incremental data. The data are processed three times a year by IQVIA, who produce and manage the pseudonymised research database (the option to identify individual patients remains

with the company but is not available to the researcher). This includes testing and checking the data and calculating data quality indicators. The data quality indicators, the acceptable mortality reporting date (AMR), the computerisation date and the Vision installation date,⁹⁴ ensure that the data made available to researchers by IQVIA are reliable and complete. See Appendix A1 for a definition of the quality indicators.

THIN data are organised by practice, followed by patient and then date meaning that all information relating to a single patient is stored together. All information is coded. For instance, clinical data in THIN are recorded using Read codes. This is a classification system for medical conditions, symptoms and referrals (Appendix A2). Information on lifestyle and preventative healthcare, immunisation and tests is available as Additional Health Data (AHD) codes. Prescribing data are organised as drug codes linked to a British National Formulary (BNF) code and include a short description and dosage information for the prescribed drug.⁸⁹

3.2.2 THIN DATA TO ADDRESS THE STUDY OBJECTIVES

Electronic patient data are usually used for cross-sectional surveys, case-control or retrospective cohort studies in epidemiological, drug safety, clinical and healthcare usage research.⁸⁸ Routine data are less often used for test accuracy studies (see section 4.3). THIN data on morbidity, prescribing and management⁹⁰ present real world evidence that are particularly useful in providing information on how the FC test is currently used in primary clinical practice. At the time the NICE recommendations were formulated, little evidence of FC testing from primary care was available.⁵⁸ The NICE evaluation defined eligibility for testing as ongoing abdominal symptoms for more than 6 weeks in patients less than 45 years of age and assumed that GPs would only test those they would otherwise have referred without testing.⁵⁸ Furthermore, they recommended referral of patients with FC levels $\geq 50\mu\text{g/g}$ and assumed that GPs would only refer patients with a positive FC test to secondary care.⁵⁸ However, early experience with FC testing in primary care reported that between 26% and 45% of patients with a negative FC test result were referred to specialist care.^{51, 55, 63} We do not know how many patients GPs consider for testing and how the FC test would impact on clinical decision making. The need for FC testing for IBD as it is moving into primary care is difficult to gauge from the existing literature because the current evidence on IBD prevalence and incidence is dated,⁹⁵ excluded a high proportion of ambiguous diagnoses,^{96, 97} or excluded cases of IBD unclassified (IBDU).⁹⁸

Therefore, I set out to investigate the uptake of FC testing into primary care, the size of the problem and the impact of testing on management decisions. My aim was to characterise FC test use in primary clinical practice to allow a critical view of the assumptions in the national recommendations.

3.3 Objectives

- To estimate the incidence and prevalence of IBD in the adult THIN population
- To describe the uptake of FC testing into primary care
- To characterise FC test use in primary care
- To determine referral rates and endoscopy rates in FC tested patients

3.4 Methods

I report the methods in three parts. Firstly, I broadly describe the THIN study in terms of the study design and study population. Secondly, I briefly explain the steps necessary to create my THIN dataset referring to the Appendix for further information. Thirdly, I detail the methods of the studies investigating IBD prevalence and incidence, FC uptake, FC test use in clinical practice and numbers of referrals and colonoscopies following FC testing.

3.4.1 DESIGN OF THE THIN STUDY

3.4.1.1 Study design

The THIN study was a retrospective cohort study of historical data collected during routine GP consultations between 1 January 2006 and 31 December 2016. The study was approved by the Independent Scientific Advisory Committee for THIN (protocol number 17THIN089, 23/10/2017).

3.4.1.2 Definition of the study population

The study population was defined by eligibility criteria at the practice level and the individual patient level. Individual practices were eligible to contribute patients to the cohort from the earliest of their AMR date and from their Vision installation date plus one year (see definitions in Appendix A1) until their last data collection date. Addition of a year following the Vision installation date was required to allow for

practices to adjust to the new computer software and to return to the same level of reporting/coding as before the software change.

The observation period was defined by the patient start and end date. Individual patients were eligible for inclusion in the study from the latest of the following dates: practice start date, patient registration date plus one year; and age 18. The year following the patient’s registration date with the THIN practice ensured that patients entering the study had at least one year of historical data available and that any of their recorded symptoms and diagnoses in the observation period were new incidents. Patients were followed-up until the first of the following dates: practice end date; patient died; patient left practice; last data collection from practice.

Figure 3.1 shows a hypothetical cohort of individual patients entering and leaving the study to help visualise the nature of the data and the different periods of follow-up available. I defined the length of follow-up for individual index dates (e.g. FC test date) and restricted the inclusion of patients to those with complete follow-up available.

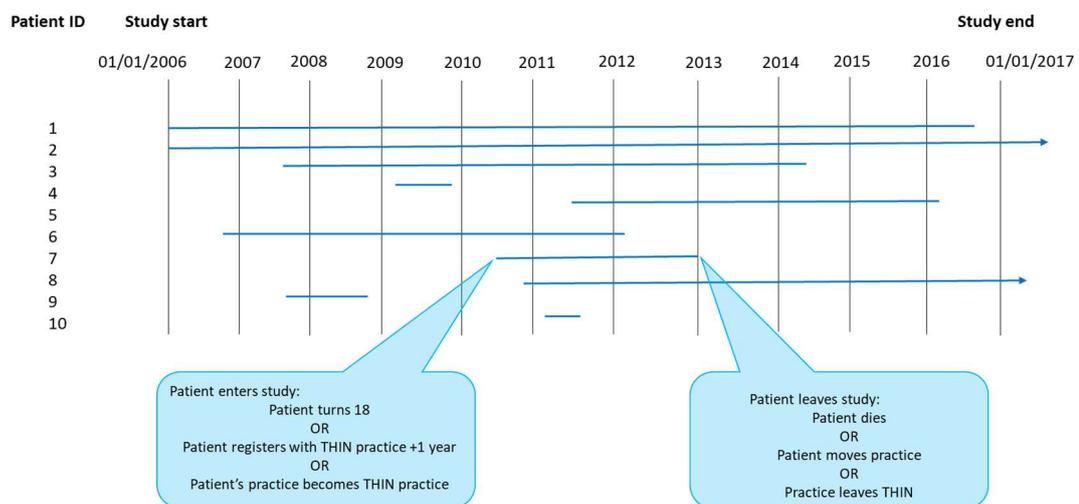


Figure 3.1 Hypothetical THIN study cohort of individual patients entering and leaving the study within the study period resulting in different length of follow-up times

3.4.2 METHODS OF CREATING THIN DATASETS

In section 3.4.1 I described the study design of the THIN study. In this section I set out the steps I followed in creating the dataset for the THIN study. These were: 1) defining the study variables, 2) creating code lists, 3) extracting data and 4) familiarisation and data management. These steps are described below.

3.4.2.1 Defining study variables

I used the research objectives in section 3.3 to devise a list of variables needed to address the research questions. Study variables fell into six main categories: 1) conditions, 2) symptoms, 3) investigations, 4) drugs, 5) referral and 6) demographics. I defined these study variables in line with my study aims and research questions in readiness to select codes to extract coded patient data from the database into a dataset. I describe how I created the code lists in section 3.4.2.2.

3.4.2.1.1 Conditions

The main condition of interest was IBD (in addition IBS and CRC were of interest for the test accuracy study in Chapter 4). IBD is a collective term of different conditions. I included Read codes for ulcerative colitis, Crohn's disease, indeterminate colitis and microscopic colitis which follow the same treatment pathway in clinical practice.

In the definition of IBS I considered any clinical record that suggested a definitive diagnosis of IBS while individual or clustered symptoms were not considered sufficiently specific for an IBS diagnosis.

Colorectal cancer was included as an alternative diagnosis to IBD and IBS following FC testing. The definition of CRC excluded benign and precancerous stages such as benign neoplasms, in-situ CRC and polyps.

Co-morbidities of IBD and IBS were considered as predictors of the two conditions. For this purpose I only considered the two most common comorbidities for IBD and IBS, respectively, namely enteropathic arthritis and asthma for IBD²⁸ and functional dyspepsia and depression for IBS.^{11, 99} The list of Read codes was not meant to be exhaustive but to detect the most relevant diseases.

The seven conditions and their definitions are described in Table 3.1.

Table 3.1 Definition of conditions for the THIN study

Condition	Definition for inclusion
IBD	Diagnosis suggestive of ulcerative colitis, Crohn's disease, indeterminate colitis or microscopic colitis
IBS	Diagnosis suggestive of IBS, symptoms including abdominal pain, abdominal distension and change in bowel habit were considered not sufficiently specific
Colorectal Cancer	Diagnosis of colorectal cancer excluding benign neoplasm, in-situ (pre-cancerous stages) and polyps
Enteropathic arthritis	Inflammatory arthritic conditions in patients with a diagnosis of IBD including reactive arthritis, rheumatoid arthritis, spondylitis and psoriatic arthritis, enteropathic arthritis, reactive arthritis caused by gut pathogens, rheumatoid arthritis, psoriatic arthritis and spondylitis and excluding Still's disease and Reiter's disease
Depression	Diagnosis of depression or low mood to include depressive symptoms with focus on diagnosis and excluding management, screening, testing, and depression as comorbidity, manic depression and psychosis excluded as separate conditions
Asthma	Any asthma diagnosis excluding tests, management and treatment of asthma
Functional dyspepsia (Non-ulcer dyspepsia)	Any diagnosis suggestive of functional dyspepsia

3.4.2.1.2 Symptoms

Symptoms of interest were abdominal complaints that would justify FC testing. Therefore, the definition of symptoms focused on undefined abdominal complaints without known unrelated causes (Table 3.2). Symptoms were based on the eligibility criteria for FC testing according to the NICE guidance⁵⁹ including a change in bowel habit, diarrhoea, constipation, bloating or abdominal pain for more than six weeks. The symptoms were used to a) describe FC tested patients and b) to define eligibility for FC testing when symptoms lasted for more than six weeks.

Table 3.2 Definitions of symptoms for the THIN study

Symptom	Definition for inclusion
Abdominal pain	Symptoms of pain in the abdomen without indication of an unrelated specific cause of the pain (e.g. hunger pain). Guarding, acute abdomen and rebound are symptoms of severe IBD and were not consider in association with FC testing as patients with these symptoms need to be referred immediately.
Bloating	Symptoms suggestive of distension of the abdomen due to gas. Flatulence was also included.
Change in bowel habit	Symptoms suggestive of a change in frequency including alternating bowel habit, rather than colour or consistency in stool. Diarrhoea and constipation were considered separately.
Constipation	Symptoms suggestive of constipation. A record of laxatives was not considered an indicator of constipation because they are also used for bowel preparation prior to bowel investigations in the patient population of interest.
Diarrhoea	Symptoms or medication suggestive of diarrhoea which were either broad / undefined or specifically relating to unexplained and non-infectious diarrhoea not caused by for instance drugs / infection / surgery

3.4.2.1.3 Investigations

Investigations of interest included the FC test as well as invasive tests such as colonoscopy and sigmoidoscopy to establish the proportion of FC tested patients who had further confirmatory testing. Any record suggesting that a patient received an FC test was considered. A record indicative of a referred patient having received or being about to receive colonoscopy or sigmoidoscopy was included. Therapeutic investigations were not included. Endoscopy was only included if the record specified that it was received for lower gastrointestinal tract investigations.

3.4.2.1.4 Drugs

I supplemented the definition of an IBD diagnosis with the record of an IBD specific prescription to increase the sensitivity of my algorithm. IBD specific drugs are mesalazine, olsalazine and balsalazide.¹⁰⁰ Sulfasalazine, prednisolone and budesonide preparations were considered IBD specific if rectal. Preparations of beclometasone needed to clearly specify use for the bowel to be included. The aim was to identify patients treated for IBD without including too many patients on

treatment for a different condition (false positives), i.e. being accurate rather than exhaustive. Therefore, the definitions were purposefully narrow and my decisions on inclusion were exclusive if in doubt. The drugs and their definitions are summarised in Table 3.3.

Table 3.3 Definitions of therapeutic drugs for the THIN study

Drug	Definition for inclusion
Mesalazine	All generic and proprietary names because drug is IBD specific
Olsalazine	All generic and proprietary names because drug is IBD specific
Balsalazide	All generic and proprietary names because active ingredient is mesalazine
Sulfasalazine	Preparations for rectal use only
Beclometasone	Tablets if preparation clearly for bowel, exclusion of inhalers, inhaler capsules, nebulisers, creams, ointments and nasal sprays
Budesonide	Preparations for rectal use only, exclusion of inhalers, nebulisers, topical treatments and nasal sprays
Prednisolone	Preparations for rectal use only, exclusion of ordinary tablets, injections, eye and ear drops

3.4.2.1.5 Referral

Referrals were defined using the standard methods described in the THIN documentation.⁹⁴ I included any referral to any speciality as well as any record of a discharge letter, outpatient appointment or hospitalisation event within six weeks of the index FC test.

3.4.2.2 Creating code lists

I used the definitions of my variables which I described in section 3.4.2.1 to establish lists of clinical codes for each study variable. The code lists were used to identify and extract coded data from the database (section 3.4.2.3).

In general, I created lists of Read codes for symptoms and diagnoses by interrogating and supplementing existing code lists.^{24, 101, 102} The aim was to identify all Read codes that can be used by GPs to record the conditions or symptoms of interest. One study validated the identification of patients diagnosed with IBD from the CPRD using GP questionnaires and reported that their code list had a positive predictive value of 92%.²⁴

To create exhaustive drug code lists I first identified all generic and proprietary names of IBD specific medications using the Prescription Cost Analysis data.¹⁰³ I then used the list of drug names to collate all codes available to GPs to record a prescribed IBD drug in the patient records.

I considered all six Read codes available to GPs to record an FC test in the code list for FC. I examined all descriptions of Read and ADH codes with the ending 'oscopy' to identify relevant codes for invasive investigations.

My code lists were reviewed by a general practitioner and we resolved discrepancies by discussion. The process of creating the code lists is described in more detail in Appendix A3 including the final code lists for all variables.

3.4.2.3 Extracting THIN data into datasets

I produced final code lists for each variable in Excel and provided a full definition of each code. A specialised THIN data analyst used the code lists to extract and collate patient data into a dataset and returned the dataset to me for cleaning, management and analysis.

3.4.2.4 Familiarisation and data management

Prior to analysis, familiarisation and data management were essential. I investigated the THIN records using summary statistics and histograms and by inspection to identify any unusual cases and devise strategies for consideration of these in the analysis. I used R version 3.6.1 (Vienna, Austria) for data management and analysis.¹⁰⁴

3.4.3 METHODS OF ANALYSIS

So far, I have covered the design of the THIN study (section 3.4.1) and the methods of creating the THIN dataset for analysis (section 3.4.2). In this section I describe the methods of using the THIN dataset to investigate the IBD prevalence and incidence, the FC uptake, the FC test use and the proportion of patients with a referral and colonoscopy record following FC testing.

3.4.3.1 IBD prevalence and incidence

I considered adult patients ≥ 18 years. I identified IBD diagnoses using the first recorded Read code for IBD or the first prescription of an IBD specific medication in the patient record.

The 12-month period prevalence of IBD was determined for the study period 2006-2016. Period prevalence was defined as new and pre-existing IBD cases during a 12-month period over the number of patients in the THIN database during that time. Confidence intervals for proportions were calculated using the Wilson procedure without a correction for continuity.

The annual incidence of IBD with exact Poisson confidence limits was determined for the same study period. Annual incidence was defined as the number of new cases of IBD during a one year period over the total time each patient was observed (person-time at risk). Patients were observed from the start of the year until, either, patients had an IBD diagnosis, had left the study or had reached the end of the observation period. IBD cases prior to 2006 and with a date before the patients' study start date were excluded from the analysis.

3.4.3.2 FC test uptake

Adult patients ≥ 18 years were included. Patients with more than one FC test were considered as separate records. I considered tests with a recorded test date after the patients' study entry date and before the patients' study end date.

I determined FC uptake as the total number of FC tested patients per 1,000 population over time. Furthermore, I determined the number of FC tests per practice defined in two ways. Firstly, I plotted the first ever FC test per THIN practice for each practice over time and, secondly, I graphed the number of FC tests ordered per 1,000 practice population by the number of practices over the study period.

Finally, I determined the proportion of patients with one or two FC tests versus patients with more than two FC tests.

3.4.3.3 Characterisation of FC test use

I considered all FC tests recorded for adult patients ≥ 18 years. I characterised all FC tests recorded after the patient's start date and before the patient's end date. I

considered the patients' demographics and their recorded symptoms within one year prior to testing in the characterisation.

Symptoms considered were non-infectious diarrhoea, bloating, constipation, abdominal pain or change in bowel habit. Eligible symptoms were defined according to NICE guidance, which stipulates eligibility for testing as chronic abdominal symptoms lasting more than six weeks.⁵⁹ For the purpose of this study patients became eligible from the first relevant recorded abdominal symptom when followed by a second recording of a relevant symptom after ≥ 6 weeks and < 3 months.

I drew a Venn diagram to summarise the proportions of patients with eligible symptoms for testing, with an FC test, with an IBD diagnosis and with any combination of those three outcomes between 2013 and 2016 using the online tool biovenn.¹⁰⁵ This analysis only included first ever recorded FC tests or first ever recorded eligible symptoms for each patient. I only included events with dates that followed the logical sequence: symptoms $<$ FC testing $<$ IBD diagnosis using mutually exclusive categories.

Finally, I recorded the IBD prevalence with 95% confidence intervals in patients eligible for FC testing for 2013-2016. I defined eligibility for FC testing in two ways; 1) patients with eligible symptoms according to the NICE guidance and 2) patients with an FC test record. An IBD diagnosis was considered when it was recorded within one year following FC testing and within two years of eligible symptoms. In the absence of clear NICE guidance on age to define eligibility for FC testing, I considered prevalence for patients < 40 , < 45 , < 50 , < 55 and < 60 years of age at time of eligibility.

Inter-practice variation in IBD prevalence was determined as median and interquartile ranges to estimate the variance of IBD prevalence between practices.

3.4.3.4 Referral and colonoscopy following FC testing

I included adult patients with at least one FC test and no prior diagnosis of IBD. To determine referrals I followed patients up from their first FC test until the earliest of the following dates: patient referred and six weeks after the index FC test date. I considered a 6-week follow-up based on evidence that IBD cases are generally referred within four weeks of FC testing,¹⁰⁶ and recommendations that FC negatives

should be reviewed within four to six weeks and referred if symptoms are ongoing.¹⁰⁷

To determine colonoscopies patients were followed up from their index FC test until the first of the following dates: GP recorded colonoscopy or one year after the index FC test date. I used one year of follow-up to account for late routine referrals and lateral referrals from gastroenterology to the surgeon.

I plotted time to referral and time to colonoscopy using Kaplan-Meier survival analysis methods for patients with positive and negative FC test outcome.

I determined proportions of FC tested patients with a referral and colonoscopy record. For this analysis, I additionally considered an IBD diagnosis as referral with colonoscopy positive. This was under the assumption that an IBD diagnosis is not usually made without a confirmatory investigation at a secondary care setting.

Finally, I determined the proportion of IBD diagnoses in referred and colonoscoped patients by FC test result. I only included patients with the complete follow-up time available after testing so as not to introduce bias against late referrals/colonoscopies in test negatives compared to test positives.

3.4.4 SUMMARISING THE METHODS

The THIN study was a retrospective cohort study from 2006-2016. I followed four steps in creating the dataset for the THIN study. These were: (1) defining the study variables, (2) creating code lists, (3) extracting data and (4) familiarisation and data management. Study variables included conditions, symptoms, drugs, referral and investigations. I defined the study population, the observation period, follow-up times and methods of analysis for each of the THIN studies which investigated the IBD prevalence and incidence, the FC uptake, the FC test use and the proportion of patients with a referral and colonoscopy record following FC testing.

3.5 Results

In this section I report the results of my studies which address the use of FC testing in primary care. This includes the prevalence and incidence of IBD, the uptake of FC into primary care clinical practice, the characterisation of FC test use and the proportion of FC tested patients with subsequent referral and colonoscopy records.

3.5.1 IBD PREVALENCE AND INCIDENCE

The first aspect of my study on FC test use in primary care considered the prevalence and incidence of IBD in the THIN population. Knowledge of the prevalence and incidence help to describe the size of the problem when implementing FC testing for adult patients in primary care.

3.5.1.1 IBD prevalence and incidence in the THIN database

I retrieved 6,965,853 records of adult patients from the THIN database. I excluded 33,730 patients who entered the study after the study period (2006-2016 see section 3.4.3.1). Furthermore, I excluded 61,129 prevalent cases from the analysis of IBD incidence. These patients had their IBD diagnosis recorded prior to the study's start date (n=50,781) or prior to the patient's start date (n=10,348). The analyses of IBD prevalence and incidence included a total of 6,932,123 and 6,870,994 patients, respectively (Figure 3.2). I did not identify any irregular cases which would have a negative impact on the study findings (see Appendix A4 for my data management decisions).

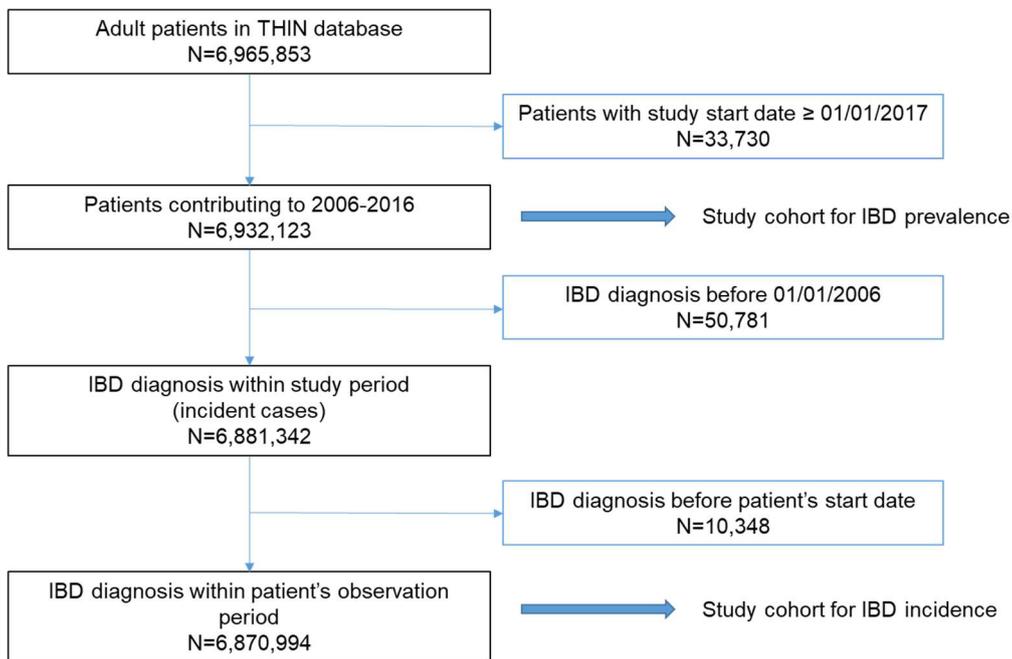


Figure 3.2 Flow diagram of inclusion and exclusion of patients into the analyses of IBD prevalence and incidence

The prevalence of IBD increased between 2006 and 2016 from 106.2 (95% CI 105.2 to 107.3) to 142.1 (95% CI 140.7 to 143.5) IBD cases per 10,000 in the adult

THIN population with an average increase of 2.96% per annum. More women than men had a recorded diagnosis of IBD (Table 3.4 and Figure 3.3).

Table 3.4 12-month period prevalence of IBD per 10,000 adult THIN population

Year	Cases	Period prevalence of IBD in adult THIN population per 10,000		
		Total (95%CI)	Male (95%CI)	Female (95%CI)
2006	37,085	106.2 (105.2; 107.3)	98.8 (97.3; 100.2)	113.5 (112.0; 115.1)
2007	39,898	110.0 (110.0; 111.1)	102.5 (101.0; 103.9)	117.4 (115.9; 119.0)
2008	43,929	113.9 (112.8; 115.0)	106.4 (105.0; 107.9)	121.2 (119.6; 122.7)
2009	46,992	118.3 (117.2; 119.3)	110.3 (108.8; 111.8)	126.0 (124.5; 127.6)
2010	48,497	122.8 (121.7; 123.9)	114.7 (113.2; 116.2)	130.7 (129.1; 132.2)
2011	49,932	126.4 (125.4; 127.6)	118.4 (116.9; 120.0)	134.2 (132.6; 135.8)
2012	52,497	129.5 (128.4; 130.6)	121.5 (119.9; 123.0)	137.2 (135.7; 138.9)
2013	52,984	132.4 (131.2; 133.5)	124.6 (123.1; 126.2)	139.8 (138.2; 141.5)
2014	51,653	135.9 (134.8; 137.1)	128.4 (126.8; 130.1)	143.1 (141.5; 144.8)
2015	48,163	139.3 (138.1; 140.5)	131.4 (129.7; 133.1)	146.9 (145.2; 148.7)
2016	40,341	142.1 (140.7; 143.5)	134.3 (132.4; 136.2)	149.6 (147.7; 151.6)
Mean (95% CI)	46,543	125.2 (118.1; 132.3)	117.4 (110.3; 124.4)	132.7 (125.6; 139.8)

CI confidence interval, IBD inflammatory bowel disease, THIN the Health Improvement Network

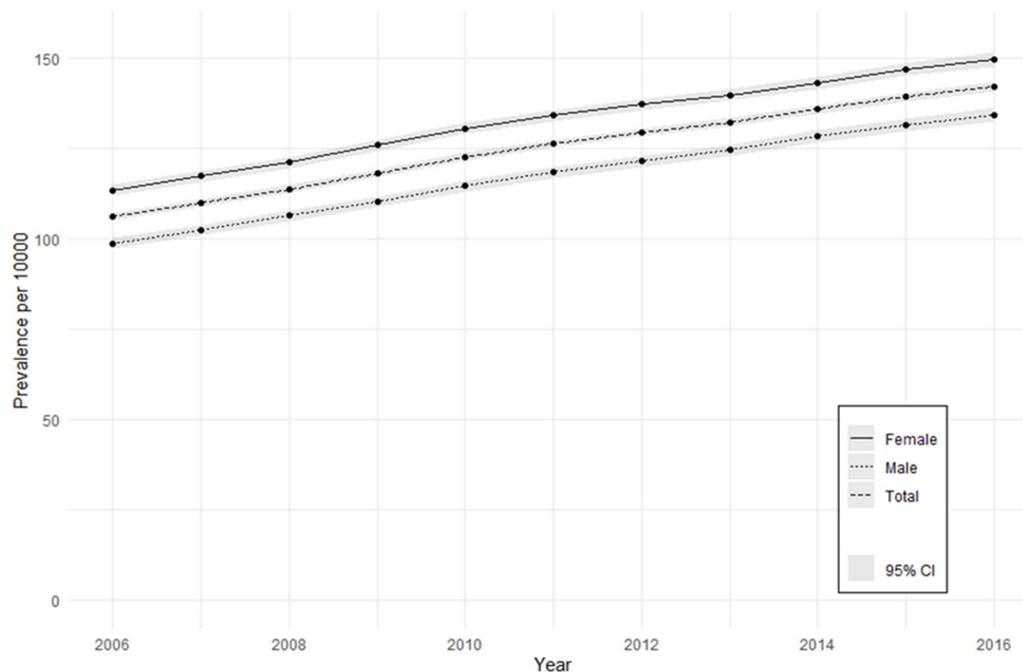


Figure 3.3 12-month period prevalence of IBD per 10,000 adult THIN population

There were 25,470 IBD incidence cases between 2006 and 2016. 4,736 (18.6%) had an IBD Read code only, 9,632 (37.8%) had a prescription of an IBD medication only and 11,102 (43.6%) had both.

Incidence of IBD in the adult THIN population varied across the years with a maximum of 76.4 (95% CI 73.6 to 79.4) per 100,000 person-years recorded in 2010 and the lowest incidence of 63.5 (95% CI 60.4 to 66.7) per 100,000 person-years recorded in 2016 (Table 3.5). The mean for the study period was 69.3 (95% CI 66.8 to 71.8) per 100,000 person-years. The incidence rate was generally higher in women than men (Figure 3.4). The drop in incidence between 2010 and 2011 may be an artefact or caused by an administrative change in coding/reporting standards. Over the most recent 5-year period, the incidence of IBD was relatively stable.

Table 3.5 Annual IBD incidence per 100,000 for the adult THIN population

Year	New cases	Annual incidence of IBD in adult THIN population per 100,000 person-years		
		Total (95%CI)	Male (95%CI)	Female (95%CI)
2006	2,295	74.7 (71.7; 77.9)	69.8 (65.6; 74.2)	79.6 (75.2; 84.1)
2007	2,361	72.0 (69.1; 75.0)	66.4 (62.4; 70.4)	77.6 (73.4; 81.9)
2008	2,473	71.6 (68.8; 74.5)	67.2 (63.4; 71.2)	76.0 (71.9; 80.2)
2009	2,687	74.7 (71.9; 77.5)	68.2 (64.4; 72.2)	81.0 (76.9; 85.2)
2010	2,740	76.5 (73.7; 79.4)	72.8 (68.9; 76.9)	80.1 (76.1; 84.4)
2011	2,387	66.6 (64.0; 69.4)	63.5 (59.8; 67.3)	69.7 (65.0; 73.6)
2012	2,399	65.5 (62.9; 68.2)	65.0(61.3; 68.8)	66.0 (62.3; 69.7)
2013	2,349	65.8 (63.2; 68.5)	65.9 (62.2; 69.8)	65.7 (62.0; 69.6)
2014	2,217	66.5 (63.8; 69.4)	61.8 (58.1; 65.7)	71.1 (67.2; 75.3)
2015	1,985	68.1 (65.1; 71.1)	62.2 (58.1; 66.4)	73.8 (69.5; 78.3)
2016	1,561	63.6 (60.5; 66.8)	61.7 (57.3; 66.3)	65.4 (61.0; 70.1)
Mean (95% CI)	2,314	69.6 (67.0; 72.2)	65.6 (63.8; 67.9)	73.3 (69.7; 76.8)

CI confidence interval, IBD inflammatory bowel disease, THIN the Health Improvement Network

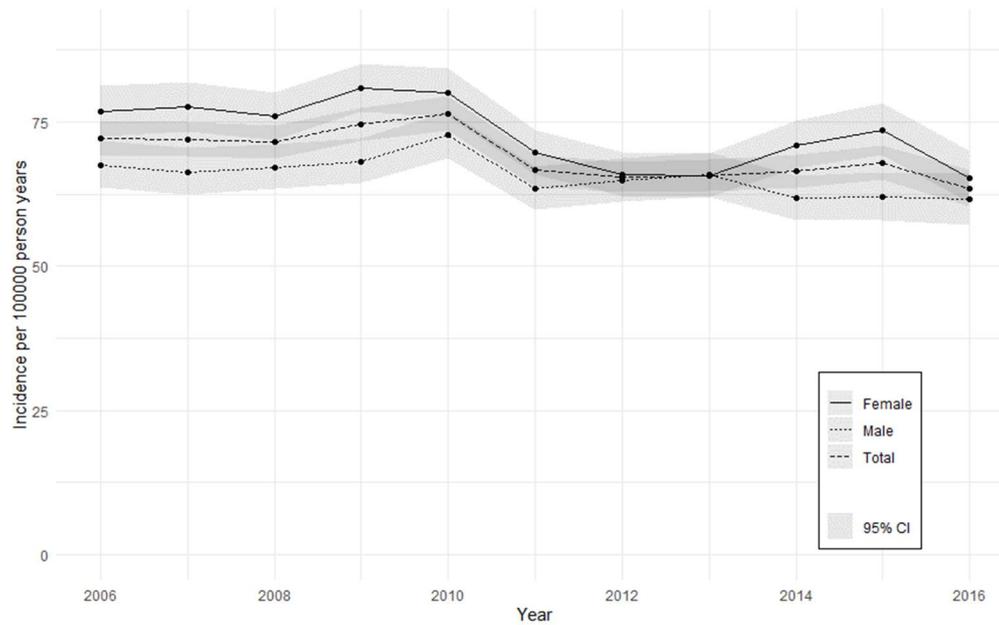


Figure 3.4 Annual IBD incidence per 100,000 person-years in the adult THIN population

3.5.1.2 Key findings from my analysis of IBD prevalence and incidence

The analyses of IBD prevalence and incidence included a total of 6,932,123 and 6,870,994 patients, respectively. The mean prevalence of IBD for the study period 2006-2016 was 125.2 (95% CI 118.1 to 132.3) per 10,000 adult patients. The prevalence of IBD increased between 2006 and 2016. The mean IBD incidence for the study period was 69.3 (95% CI 66.8 to 71.8) per 100,000 person-years. Over the most recent 5-year period, the incidence of IBD was relatively stable.

3.5.2 FC TEST UPTAKE

The second aspect I considered was uptake of FC testing into clinical practice including all FC tests recorded in the THIN database. More than one FC test could belong to the same patient. Investigating the uptake of FC testing not only provides an understanding of the extent the test is established in primary care but also indicates the impact of NICE guidance on test use.

3.5.2.1 FC test uptake into clinical practice according to THIN

There were 22,299 FC tests recorded in the THIN population of 6,965,853 adult patients (6,969,041 patient records) between 2006 and 2016. Following exclusion of records of patients with a study start date after 31/12/2016, the study included 6,935,283 records (Figure 3.5) with 22,137 FC tests. 2,297/22,127 tests were

recorded outside the study period (either before the patient's study start date or after 31/12/2016). I removed the test variable from these records but did not exclude the records from the study to allow them to be counted in the denominator.

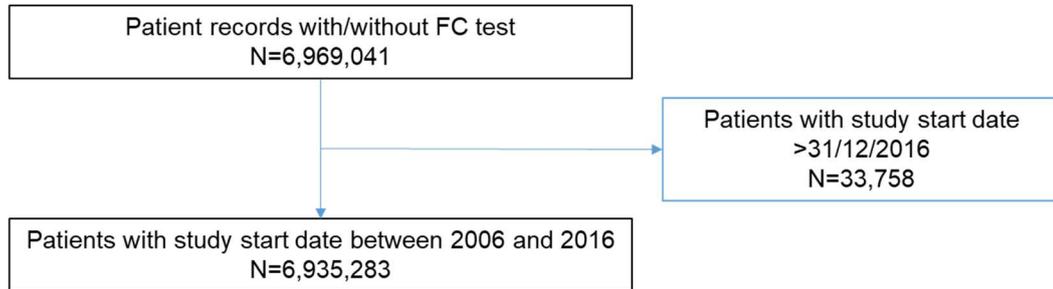


Figure 3.5 Overview of inclusion and exclusion of patient records with and without FC tests into the study of FC uptake

Figure 3.6 depicts FC tested patients per 1,000 population over time. The first FC tests were documented in 2009 with a noticeable increase in FC test use in 2013. FC test use was still increasing at the end of the study period with a slightly lesser incline from 2015 to 2016.

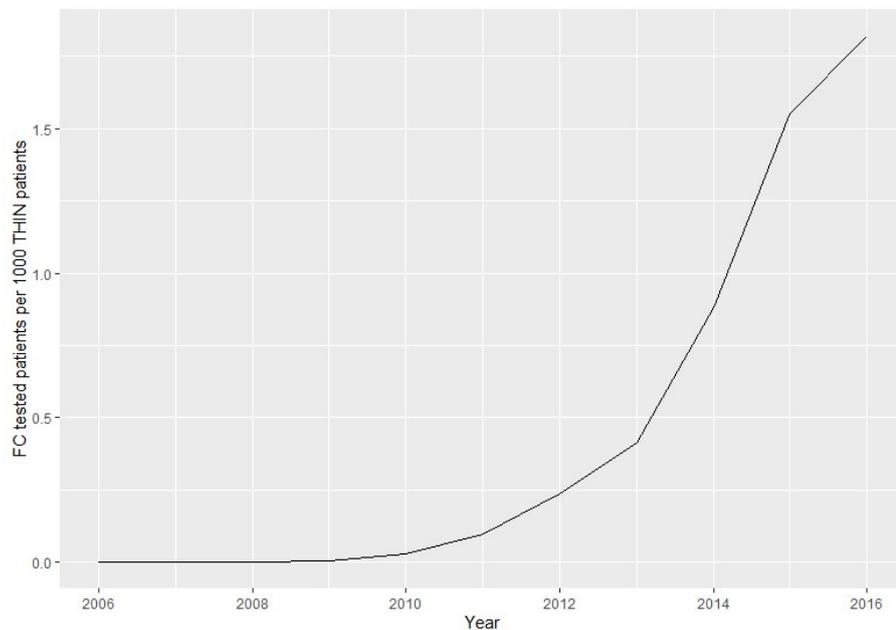


Figure 3.6 FC tested patients per 1,000 THIN population by year

Uptake by practice revealed that the number of practices recording their first FC test increased steeply between 2009 and 2012, peaked in 2014 with fewer practices

starting FC testing in 2015 and 2016 (Figure 3.7). There was no noticeable increase in practices starting FC testing in 2013. By 2016 66.8% (493/738) THIN general practices had started using FC testing. The vast majority of practices used less than five FC tests per 1,000 patients in the study period where the average number of patients per practice was 9,390. A small number of practices could be described as high users of testing according to Figure 3.8 with 40 to 58 FC tests/1,000 patients.

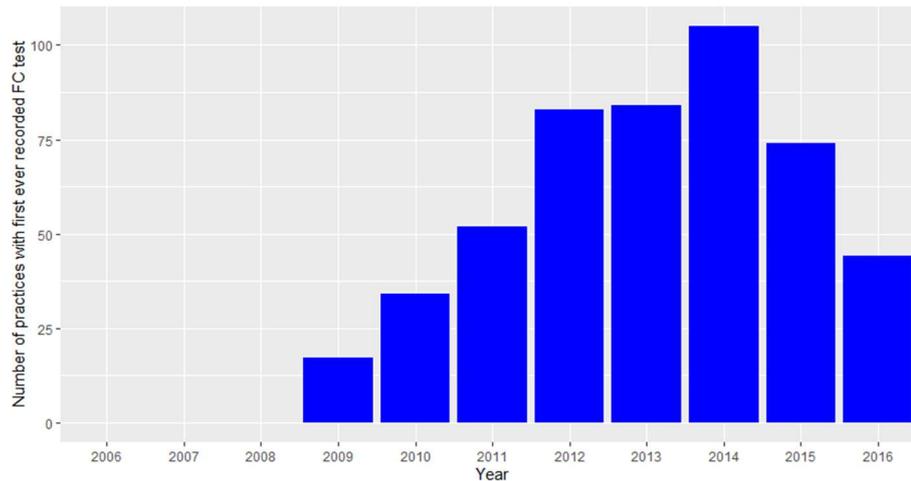


Figure 3.7 Number of practices with first ever FC test record by year

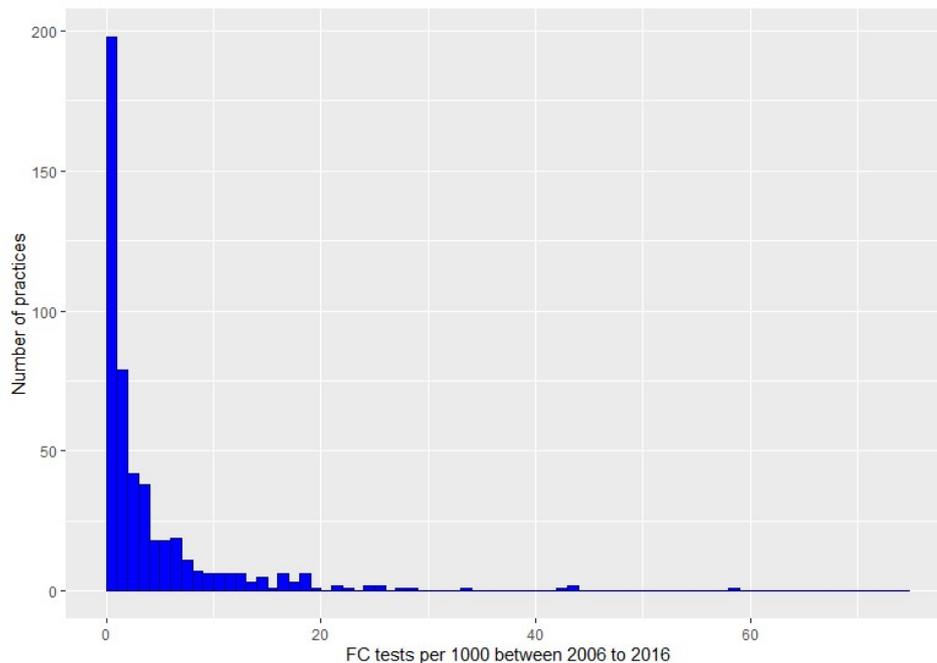


Figure 3.8 Number of FC tests recorded per 1,000 practice population by number of practices during 2006-2016
 400/17,027 (2.3%) FC tested patients had more than two FC tests in the study period which might be indicative of tests being used for monitoring rather than

diagnosis. The number of tests per patient in the study period ranged from one to 13 (median = 1, IQR = 0) (Table 3.6).

Table 3.6 Number of tests per patient between 2009 and 2016

Number of FC tests in study period	Number of patients
1	14,807
2	1,820
3	285
4	79
5	21
6	3
7	6
8	2
9	2
10	1
13	1

3.5.2.2 Key findings from my analysis of FC test uptake

There were 19,830 FC tests recorded in the study period. The first FC tests were documented in 2009 and FC test use was still increasing in 2016. By 2016 66.8% THIN general practices had started using FC testing. The majority of practices used less than five FC tests per 1,000 patients in the study period.

3.5.3 CHARACTERISATION OF FC TEST USE

In this third section of results I characterise FC test use in three ways. Firstly, I detail the patient characteristics for each FC test recorded in the THIN database.

Secondly, I summarise the proportions of patients with eligible symptoms for testing, with an FC test, with an IBD diagnosis and with any combination of those three outcomes. Finally, I describe the IBD prevalence in eligible patients for FC testing.

The findings provide some insight on who GPs test and whether the tested population is consistent with the NICE recommendations.

3.5.3.1 Characteristics of FC tests in THIN

Table 3.7 provides a summary of the demographics and clinical characteristics of FC tested patients per recorded FC test from 2009 to 2016. The data for the years 2009 to 2011 were sparse and may be atypical. Between 2012 and 2016, the mean age of the FC tested population was about 44 years of age of whom over 60% were females. About 20% of tests were in patients aged 60 years and over in whom testing has not been recommended.⁵⁵ Tested patients were generally white, came

from all socioeconomic backgrounds according to the Townsend score and had a body mass index (BMI) that fell in the overweight range. Abdominal pain and diarrhoea were the most commonly recorded abdominal symptoms within one year prior to an FC test. Bloating, constipation and change in bowel habit were reported infrequently. Overall, reporting of abdominal symptoms prior to FC testing decreased from 2011 to 2016. In 2016, only just over half of the tested patients had any abdominal symptoms recorded before testing. In particular, only 7.8% of FC test records were preceded by a record of symptoms eligible for FC testing according to NICE recommendations (see section 3.4.3.3 for the definition of NICE eligibility).

Table 3.7 Demographic and clinical characteristics of FC tested patients per recorded FC test between 2009 and 2016

	2009	2010	2011	2012	2013	2014	2015	2016
FC tests (n)	23	121	426	1,111	2,021	3,878	6,199	6,051
Age (mean (SD)) at FC test date	47.3 (17.6)	46.4 (16.9)	43.6 (16.8)	44.4 (16.6)	44.5 (16.7)	43.7 (16.3)	43.7 (16.5)	44.5 (16.9)
Age in years n (%)								
18-29	4 (17.4)	24 (19.8)	111 (26.1)	253 (22.8)	474 (23.5)	912 (23.5)	1,445 (23.3)	1,351 (22.3)
30-39	3 (13.0)	21 (17.4)	83 (19.5)	226 (20.3)	373 (18.5)	803 (20.7)	1,326 (21.4)	1,290 (21.3)
40-49	5 (21.7)	24 (19.8)	81 (19.0)	224 (20.2)	452 (22.4)	817 (21.1)	1,262 (20.4)	1,215 (20.1)
50-59	7 (30.4)	19 (15.7)	66 (15.5)	199 (17.9)	322 (15.9)	643 (16.6)	1,036 (16.7)	942 (15.6)
60-69	1 (4.3)	19 (15.7)	52 (12.2)	104 (9.4)	215 (10.6)	419 (10.8)	647 (10.4)	668 (11.0)
>=70	3 (13.0)	14 (11.6)	33 (7.7)	105 (9.5)	185 (9.2)	284 (7.3)	483 (7.8)	585 (9.7)
Sex (female (%))	20 (87.0)	85 (70.2)	281 (66.0)	689 (62.0)	1,236 (61.2)	2,458 (63.4)	3,938 (63.5)	3,850 (63.6)
BMI (mean (SD))	27.4 (6.8)	26.7 (5.4)	26.0 (6.0)	26.5 (5.9)	26.8 (6.2)	26.6 (6.2)	27.0 (6.2)	27.0 (6.1)
Not reported n (%)	1 (4.3)	9 (7.4)	31 (7.3)	104 (9.4)	224 (11.1)	378 (9.7)	672 (10.8)	697 (11.5)
Ethnicity n (%)								
White	13 (56.5)	75 (62.0)	233 (54.7)	618 (55.6)	1,186 (58.7)	2,291 (59.5)	3,724 (60.1)	3,395 (56.1)
Asian	0 (0.0)	2 (1.7)	4 (0.9)	26 (2.3)	38 (1.9)	88 (2.3)	133 (2.1)	165 (2.7)
Black	1 (4.3)	2 (1.7)	4 (0.9)	12 (1.1)	13 (6.4)	35 (0.9)	77 (1.2)	99 (1.6)
Mixed	0 (0.0)	2 (1.7)	0 (0.0)	1 (0.0)	7 (0.3)	20 (0.5)	37 (0.6)	44 (0.7)
Other	0 (0.0)	1 (0.8)	2 (0.5)	3 (0.3)	9 (0.4)	18 (0.5)	36 (0.6)	37 (0.6)
Not reported	9 (39.1)	39 (32.2)	183 (43.0)	451 (40.6)	768 (38.0)	1,426 (36.8)	2,192 (35.4)	2,311 (38.2)
Townsend score n (%)								
1 (least deprived)	0 (0.0)	19 (15.7)	53 (12.4)	133 (12.0)	347 (17.1)	693 (17.9)	1,271 (20.5)	1,213 (20.0)
2	5 (21.7)	22 (18.2)	71 (16.7)	191 (17.2)	310 (15.3)	756 (19.5)	1,117 (18.0)	1,105 (18.3)
3	2 (8.7)	25 (20.7)	58 (13.6)	204 (18.4)	369 (18.3)	718 (18.5)	1,155 (18.6)	1,154 (19.1)
4	7 (30.4)	16 (13.2)	81 (19.0)	190 (17.1)	434 (21.5)	686 (17.7)	1,163 (18.8)	1,069 (17.7)
5 (most deprived)	7 (30.4)	30 (24.8)	128 (30.0)	276 (24.8)	393 (19.4)	614 (15.8)	821 (13.2)	776 (12.8)
Missing	2 (8.7)	9 (7.4)	35 (8.2)	117 (10.5)	168 (8.3)	411 (10.6)	672 (10.8)	734 (12.1)
Symptoms recorded within 1 year prior to FC test n (%)								
Any	21 (91.3)	114 (94.2)	376 (88.3)	954 (85.9)	1,265 (62.6)	2,390 (61.6)	3,732 (60.2)	3,409 (56.3)
NICE eligible	10 (43.5)	49 (40.5)	140 (32.9)	316 (28.4)	249 (12.3)	462 (11.9)	588 (9.5)	473 (7.8)
Abdominal pain	8 (34.8)	54 (44.6)	163 (38.3)	420 (37.8)	566 (28.0)	1,084 (28.0)	1,638 (26.4)	1,568 (25.9)
Bloating	0 (0.0)	5 (4.1)	26 (6.1)	58 (5.2)	53 (2.6)	163 (4.2)	258 (4.1)	198 (3.3)
Change in bowel habit	1 (4.3)	5 (4.1)	26 (6.1)	88 (7.9)	90 (4.5)	219 (5.6)	367 (5.9)	301 (5.0)
Constipation	0 (0.0)	9 (7.4)	20 (4.7)	55 (5.0)	43 (2.1)	82 (2.1)	126 (2.0)	119 (2.0)
Diarrhoea	15 (65.2)	66 (54.5)	194 (45.5)	463 (41.7)	566 (28.0)	941 (24.2)	1,448 (23.4)	1,324 (21.9)
None	2 (8.7)	7 (5.8)	50 (11.7)	157 (14.1)	756 (37.4)	1,488 (38.4)	2,467 (39.8)	2,642 (43.7)

3.5.3.2 Proportions of patients with eligible symptoms, FC testing and IBD diagnosis

Figure 3.9 describes the size of the dataset of patients with NICE eligible symptoms for FC testing, with an FC test record or with an IBD diagnosis between 2013 and 2016. Of these 55,477 had eligible symptoms as defined by NICE guidance, 13,877 had an FC test and 7,640 had an IBD diagnosis recorded. Among these were 2,974 patients with a combination of these events.

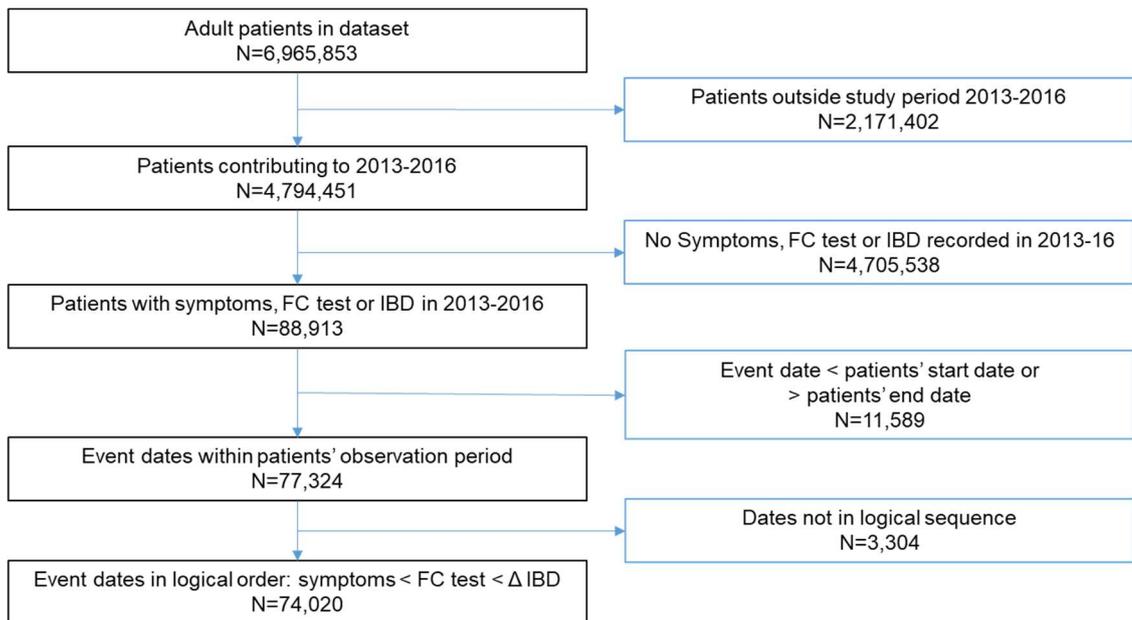


Figure 3.9 Overview of inclusion and exclusion to create a dataset of patients with NICE eligible symptoms, FC tests and/or IBD diagnosis

The Venn diagram in Figure 3.10 shows the proportional relationship of the three events in the population between 2013 and 2016. The number of patients with eligible symptoms followed by an FC test and an IBD diagnosis was small (79/74,020). Only 1,720/55,477 (3.1%) patients with eligible symptoms received an FC test. Furthermore, only 500/7,640 (6.5%) patients with an IBD diagnosis had a prior FC test. The proportion of FC tests taken in patients without eligible symptoms and without subsequent IBD diagnosis was much larger. Clearly patients following the anticipated NICE pathway (symptoms followed by FC test followed by IBD diagnosis) are in a minority. Restricting the analysis to the year 2016 (including a total of 15,151 patients) did not show a change in the pattern (Figure 3.10).

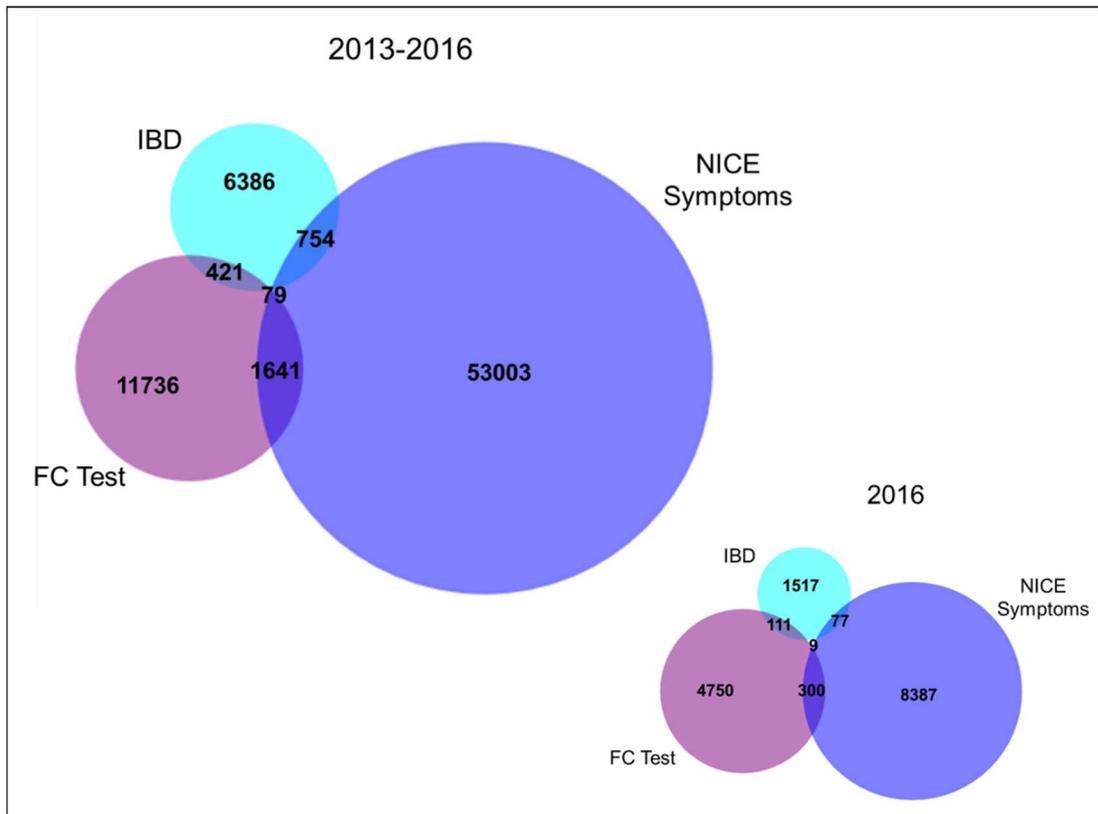


Figure 3.10 Number of patients with NICE eligible symptoms for FC testing, a record of FC testing and/or an IBD diagnosis for a four year period and for 2016 only

3.5.3.3 IBD prevalence in eligible patients

The IBD prevalence in 55,477 patients with eligible symptoms recorded between 2013 and 2016 was 1.4% (95% CI 1.3 to 1.5%). The IBD prevalence in 13,877 FC tested patients was 3.4% (95% CI 3.1 to 3.7%) (Table 3.8). Adding age as an eligibility criterion showed that prevalence was similar using age cut-offs from 40 years to 60 years with point estimates slightly decreasing. 27.8% of IBD diagnoses in FC tested patients were in patients 50 years and over.

Differences in IBD prevalence in eligible patients by practice were negligible with medians of 0 showing no variation across practices. I have chosen not to report the results.

Table 3.8 IBD prevalence in eligible patients for FC testing defined as patients with 1) a recorded period of NICE eligible symptoms and 2) a recorded FC test for the study period 2013-2016

	Eligible symptoms* N	IBD within 2 years following record of eligible symptoms N	IBD prevalence in NICE eligible patients (95% CI)	FC tests** N	IBD within 1 year following FC test N	IBD prevalence in FC tested patients (95% CI)
All ages	55,477	794	0.014 (0.013; 0.015)	13,877	468	0.034 (0.031; 0.037)
< 40 years at testing / symptoms recorded	14,708	260	0.018 (0.016; 0.020)	6,352	296	0.040 (0.036; 0.046)
< 45 years at testing / symptoms recorded	18,419	318	0.017 (0.015; 0.019)	7,787	287	0.038 (0.034; 0.043)
< 50 years at testing / symptoms recorded	22,488	366	0.016 (0.015; 0.018)	9,266	338	0.036 (0.033; 0.041)
< 55 years at testing / symptoms recorded	26,805	429	0.016 (0.015; 0.018)	10,420	379	0.036 (0.033; 0.040)
< 60 years at testing / symptoms recorded	30,879	476	0.015 (0.014; 0.017)	11,435	397	0.035 (0.031; 0.038)

*Eligible symptoms without prior record of IBD diagnosis or FC test

**FC tests without prior record of IBD diagnosis

CI confidence interval, FC faecal calprotectin, IBD inflammatory bowel disease

3.5.3.4 Key findings from my analysis of the characterisation of FC test use

About 20% of patients with a recorded FC test were aged 60 years or over. Only 7.8% of FC test records were preceded by a record of symptoms eligible for FC testing according to NICE recommendations. Furthermore, only 3.1% of patients with eligible symptoms received an FC test. 6.5% of IBD patients had an FC test recorded prior to their diagnosis. Overall, there was only a small overlap in patients with symptoms, FC test and IBD diagnosis. The IBD prevalence in FC tested patients was higher than in those with eligible symptoms. A number of patients were still diagnosed with IBD in their 50s and older.

3.5.4 REFERRAL AND COLONOSCOPY FOLLOWING FC TESTING

In this final section of results I report the proportions of FC tested patients with a referral or colonoscopy record to explore the impact that FC testing has on patient management and to contrast this with the modelled behaviour in national guidance.

3.5.4.1 Time to referral and colonoscopy

This study included 7,084 first FC tests with results (section 4.3.4.1.2) recorded in the THIN database between 2009 and 2016. Figure 3.11 shows the time from FC testing to referral by FC test result. It clearly shows that due to the definition of referral, which considers a referral to any speciality, all FC tested patients were eventually referred. However, patients with a test positive outcome were referred sooner. Furthermore, by six weeks following FC testing nearly 50% of test negatives and about 65% of test positives had been referred.

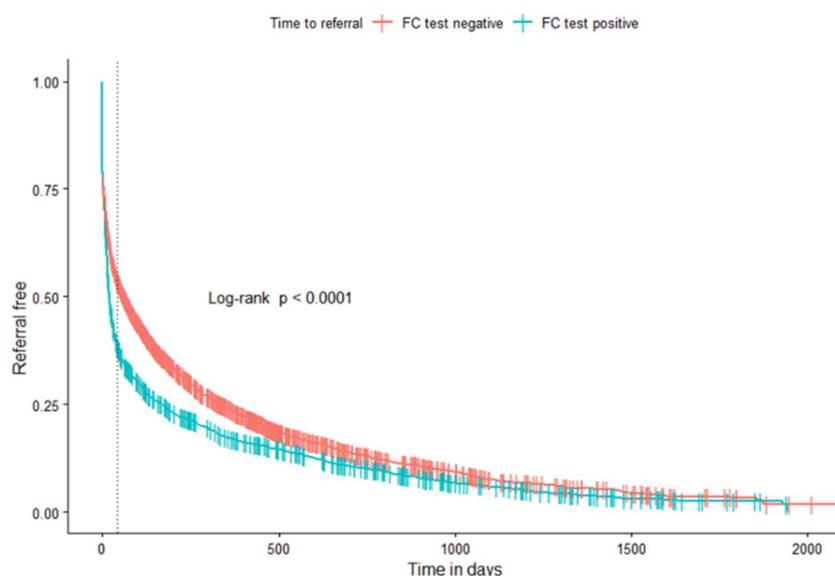


Figure 3.11 Time from FC testing to referral by FC test result (the dotted line marks six weeks following FC testing)

Figure 3.12 depicts the time from FC testing to a recorded colonoscopy and reveals that test positives had a colonoscopy sooner and more frequently than test negative patients. By one year after the FC test the vast majority of colonoscopies had happened, with very few recordings after one year.

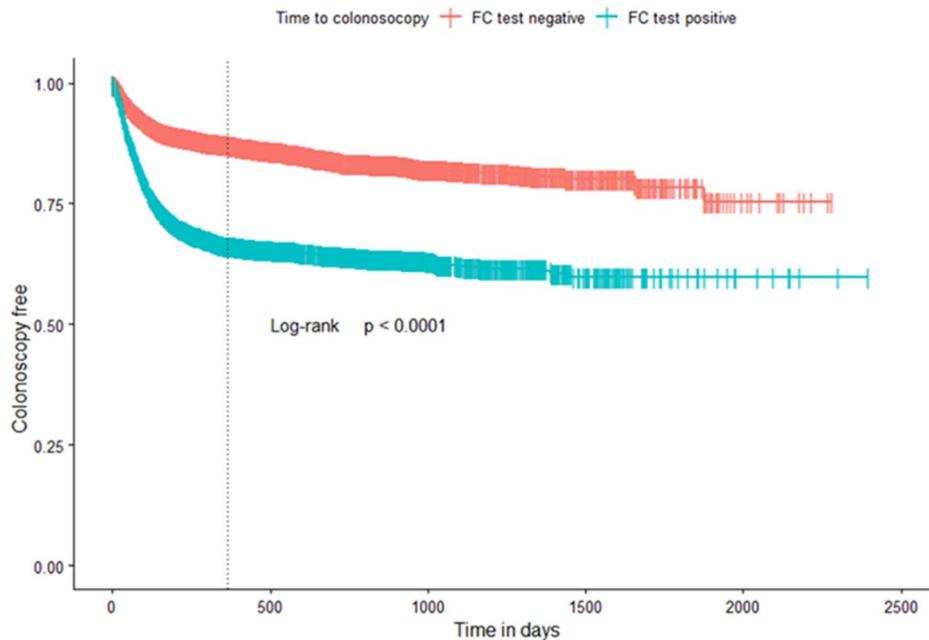


Figure 3.12 Time from FC testing to colonoscopy by FC test result (the dotted line marks one year following FC testing)

3.5.4.2 Referral within six weeks and colonoscopy within one year of FC testing

6,916 patients with FC tests had at least six weeks follow-up available following FC testing (Figure 3.13). Of these, 2,718 (39.3%) had a positive test result, while 4,198 (60.7%) had a negative result. 63.5% (1,725/2,718) of patients with a positive FC test result were referred within six weeks of FC testing and 47.6% (1,998/4,198) with a negative test result were referred.

4,792 patients with FC tests had at least one year follow-up available after FC testing (Figure 3.13). Of these, 1,987 had a positive FC test and 801 (40.3%) had a subsequent colonoscopy within one year of the test. 381/2,805 (13.6%) patients with a negative test results had a colonoscopy recorded.

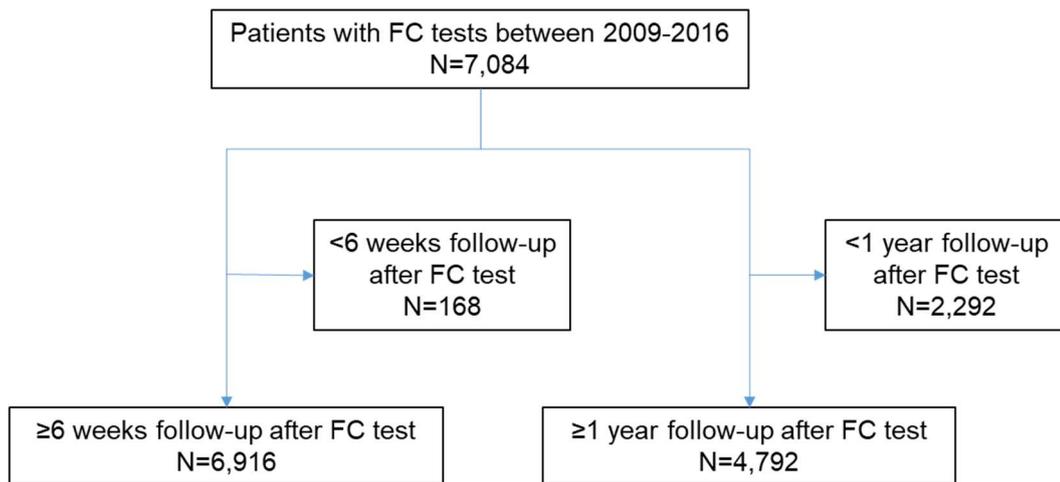


Figure 3.13 Patients with sufficient follow-up available after FC testing for consideration of referral within six weeks and colonoscopy within one year following FC testing

Figure 3.14 shows the proportion of IBD diagnoses in referred and colonoscoped patients. IBD diagnoses are shown by FC test result for 4,792 FC tested patients who had a minimum of one year follow-up available following testing. Nearly 50% (1,337/2,805) of FC test negative patients were referred and/or received a colonoscopy in the defined time periods with only a small yield of IBD diagnoses. The cases with a colonoscopy but without referral were due to the different definitions of follow-up for referral and colonoscopy events.

Table 3.9 presents my findings within the context of published evidence from primary care studies. It shows some variability but no inconsistencies in the reported estimates. My study showed a higher proportion of FC positive tests than the published studies (1,987/4,792 vs 148/410, $z=2.1$, $p=0.034$), possibly, because numeric results were recorded more frequently for positive than negative tests. However, subsequent proportions of referrals, colonoscopies and IBD diagnoses were within published figures. Studies which used a greater FC threshold reported a lower proportion of FC positive results.

3.5.4.3 Key findings from my analysis of referral and colonoscopy following FC testing

Patients with a positive FC test result were referred and received a colonoscopy sooner than patients with a negative FC test. Nearly 50% of FC test negative patients were referred and/or received a colonoscopy within one year of FC testing.

In these, only a small number of IBD cases were detected. My findings were in line with evidence from published studies in primary care.

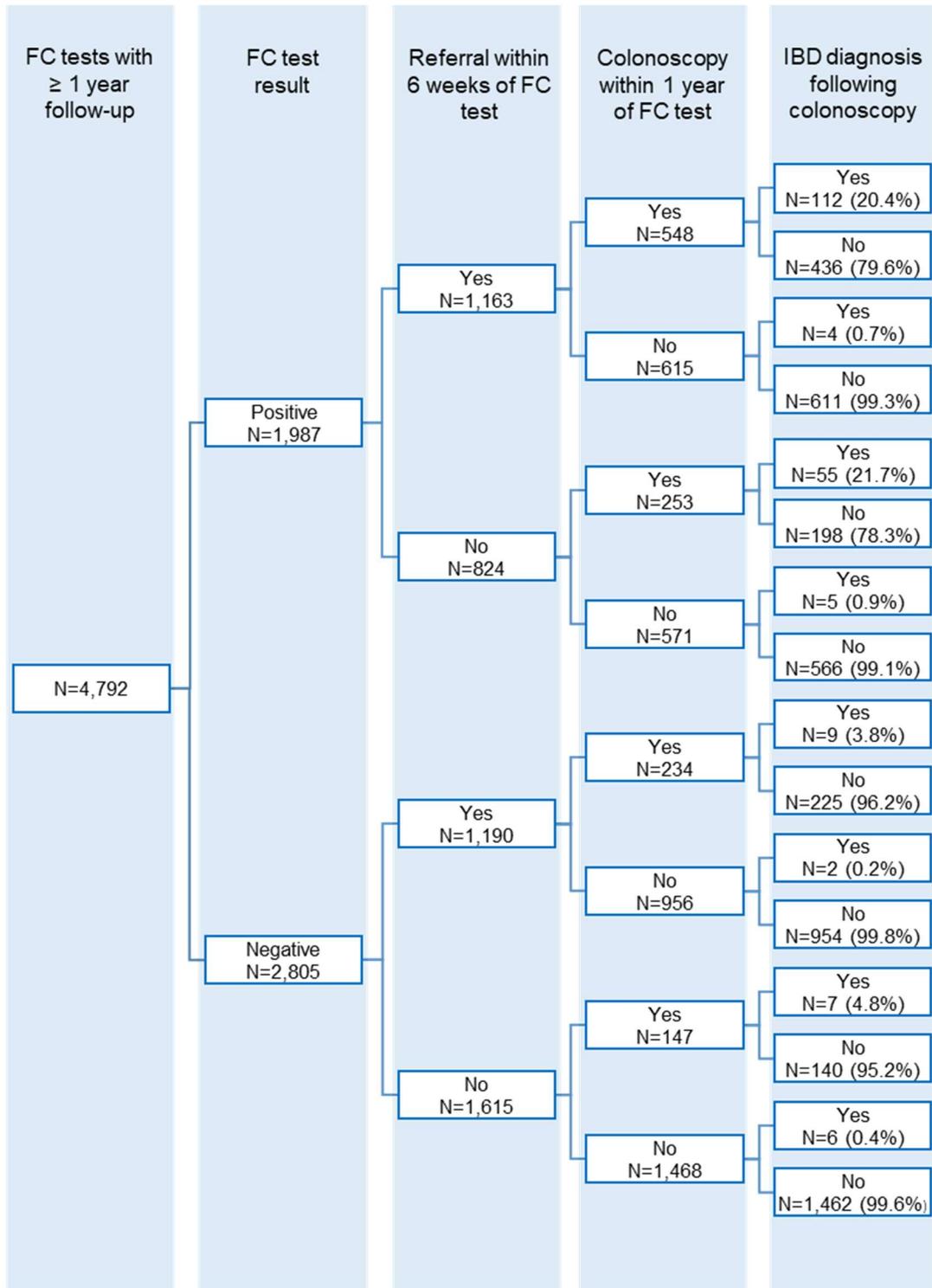


Figure 3.14 Proportions of IBD diagnoses in referred and colonoscoped patients by FC test result

Table 3.9 Proportions of referred and colonoscoped FC tested patients in published studies compared to the present study of FC tested patients with at least 1 year follow-up

Study reference	FC tested patients	FC threshold in µg/g	FC positive (% of tested patients)	Referred (% of tested patients)	Colonoscoped (% of tested patients)	IBD (% of tested patients)
My study	4,792	50	1,987 (41.5)	2,353 (49.1)	1,182 (24.7)	214 (4.5)
Hogberg 2017 ⁶²	373	50	109 (29.2)	NR	157 (42.1)	10 (2.7)
Boyed 2016 ⁶¹	424	50	140 (33.0)	NR	NR	23 (5.4)
Conroy 2017 ⁶³	410	50	148 (36.1)	196 (47.8)	146 (35.6)	11 (2.7)
Pavlidis 2013 ⁵¹	962	50	276 (28.7)	406 (42.2)	368 (38.3)	42 (4.4)
Walker 2018 ⁶⁴	789	100	132 (16.7)	405 (51.3)	299 (37.9)	50 (6.3)
Turvill 2016 ⁵⁵	262	100	64 (24.4)	108 (41.2)	66 (25.2)	9 (3.4)
Turvill 2018 ⁶⁵	951	100	143 (15.0)	305 (32.1)	211 (22.2)	52 (5.5)

FC faecal calprotectin, IBD inflammatory bowel disease

3.6 Discussion

3.6.1 SUMMARY OF STUDY FINDINGS

My analyses on the use of FC testing in primary care revealed that uptake of FC testing has been slow. Four years after the publication of the NICE guidance on faecal calprotectin, only 64% of THIN practices had started using FC testing and the frequency of testing varied widely across practices. About 20% of tests were carried out in patients aged 60 years and over in whom testing has not been recommended.⁵⁸ Moreover, only 7.8% of FC test records were preceded by symptoms eligible for FC testing according to NICE recommendations.⁵⁸ Overall, there was little overlap between patients with eligible symptoms, patients with FC testing and patients with an IBD diagnosis. Only 3.1% of patients with eligible symptoms received FC testing. Furthermore, only 6.5% of IBD patients had a record of an FC test prior to their diagnosis. Of the tested population, 42% of patients received a positive test result. Surprisingly, nearly 50% of FC test negative patients were referred and/or received a colonoscopy which was not indicated by the FC test result. However, referral rates and colonoscopy rates in FC tested patients were broadly in line with published figures which may suggest a lack of understanding and consensus among GPs on who to test and how to act on the test result. My analyses contribute to our understanding of how FC testing has been introduced into primary care clinical practice. The findings indicate non-compliance with NICE guidance and provide little insight into who GPs test.

In my discussion of the study findings, I first highlight the strengths and limitations of routine data from primary care for my research purposes. I then discuss the individual studies in the context of published evidence.

3.6.2 STRENGTHS AND LIMITATIONS OF PRIMARY CARE ELECTRONIC HEALTH CARE RECORDS FOR RESEARCH

The THIN database is a rich source of routine electronic health care records of patients managed in primary care and is particularly useful for the study of real world problems. The study population was large and covered nearly 50% of all UK CCGs⁸⁸ meaning that findings are generalisable to UK primary care in general.

Limitations that might have affected the research are linked to characteristics of routine data.

3.6.2.1 Defining follow-up times

I carefully considered the study objectives in the definitions of the study variables and I agreed code lists with a general practitioner. However, the time periods between index dates and events were defined somewhat arbitrarily in the absence of clear guidance. I considered published evidence and clinical advice in the definition of follow-up times. However, follow-up periods were not robust because of variation in clinical practice and possible delayed coding. Defining follow-up periods requires a balance of ensuring index dates and events are related whilst avoiding missing potentially related events. I explored different follow-up times in the analysis of time from FC test to referral/colonoscopy to explore the potential impact of uncertain follow-up times.

3.6.2.2 Study inclusion criteria

The criterion “registration date plus one year” to assess patients’ eligibility for study inclusion avoided the systematic over-reporting of incidence rates in the first year of follow-up for newly registered patients.⁹¹ It also prevented the double counting of prevalent cases when patients transferred from one THIN practice to another. This holds true when estimating incidence and prevalence by year. However, in section 3.5.3.3 I estimated the IBD prevalence in eligible patients for FC testing for 2013-2016. I may have double counted patients with an IBD diagnosis following FC testing who transferred from one THIN practice to another in this time period. It is impossible to track patients and identify those who transferred because data are anonymised. Consequently, the size of the issue is not known (a study investigating CRPD data linked with Hospital Episode Statistics (HES) reported 2% of patients with one HES ID and multiple CPRD identities suggesting they were the same patient¹⁰⁸). However, THIN estimates were similar to published figures in a comparison with primary care studies.

3.6.2.3 Missing data

Routine data are limited by what has been recorded during patient consultations. In the THIN data, 30% of FC tests had no result recorded. Missing data reduced the available data for analysis and potentially biased my studies if missing results were systematically different. The proportion of abnormal test results was slightly higher in my study than in published primary care studies. This might be due to a slight over-reporting of test positive results compared to negative results and may have affected the proportions reported for the test positive rate, referral and colonoscopy

rates. However, I expect this problem of preferential recording of positive results to be small as test results are increasingly recorded electronically.^{109, 110} Furthermore, my estimates of referral and colonoscopy were in agreement with published figures from several primary care studies. Missing test results may have had little impact on my overall study conclusions.

In contrast to missing FC test results, I do not know the extent of potentially missing IBD diagnoses either due to incorrect coding, missed coding or recording as free-text. This might have led to an underestimation of IBD incidence and prevalence. I included a record of an IBD specific medication in the definition of an IBD diagnosis to mitigate the effect. This may explain my higher figures for IBD incidence and prevalence when compared to a recent study which only included patients with two IBD Read codes recorded or one IBD Read code and an IBD drug code.⁹⁷ Without a validation of my code list to identify IBD patients, I do not know the sensitivity and specificity of the codes within the list. One study, which validated their codes to identify IBD in the CPRD, reported a positive predictive value of 92% but no negative predictive value.²⁴

To mitigate against missing referrals,^{108, 109} I used the recommended definition of a referral.⁹⁴ This included outgoing referrals as well as incoming patients from any specialist. I used the time window of six weeks in which the referrals were considered to ensure the referral was relevant to the index FC test. McBride et al. (2010) used a more conservative window of two weeks which considered only outgoing referrals.¹¹¹ Their aim was to ensure that the index symptom was associated with the observed referral. In contrast, my aim was to explore whether test negative patients were referred. At six weeks nearly 50% of FC negative patients had a referral. The 6-week window, therefore, provided a conservative estimate of referrals of test negatives in the THIN data.

Finally, secondary care investigations such as colonoscopy cannot be fully accounted for using primary care data. I considered a record of an IBD diagnosis as colonoscopy positive to mitigate against the effect. This was under the assumption that IBD diagnoses are not usually made without confirmatory testing of type, location and severity using secondary care invasive testing which determines treatment choice. I was unable to confirm to what extent this was successful. In 2011 32.2 colonoscopies per 100,000 population were recorded in the UK.¹¹² Only 7% resulted in a diagnosis of IBD since the majority of bowel investigations were

undertaken for suspicion of bowel cancer.¹¹² Therefore, the figures have little relevance to my study for comparison of colonoscopies rates in FC tested patients to gauge how many colonoscopies I may have missed. I showed that colonoscopies in FC tested patients ranged quite widely in test accuracy studies from primary care (Table 3.9). My estimate was within that range. Secondary care data such as HES may have been useful to explore the completeness of colonoscopies reported. However, linkage to HES records is only available for 17% of patients in the THIN database.⁹⁴ It is uncertain what I may have gained from considering available HES data. Consequently, I may have underestimated colonoscopies in FC tested patients in my study.

3.6.2.4 Misclassification

I was interested in the broad category of inflammatory bowel disease with no concern over the sub-category, severity or location of disease. Therefore, misclassification through miscoding of ulcerative colitis as Crohn's disease and vice versa, or by using higher order codes rather than disease specific codes would not have affected my analysis.

3.6.2.5 Definition of study variables

Finally, a limitation of the study was the definition of the NICE eligible symptoms for FC testing. The recommendations in the diagnostics guidance (DG11) very broadly recommend FC testing "in adults with recent onset lower gastrointestinal symptoms for whom specialist assessment is being considered".⁵⁹ In the absence of specific symptoms and a definition of 'recent onset' I referred to the underlying review and health technology assessment for guidance. It defined chronic lower gastrointestinal symptoms as abdominal pain, bloating or change in bowel habit persisting for at least 6-8 weeks.⁵⁸ Operationalising the definition proved difficult. I defined chronic symptoms as having two instances of a record of any gastrointestinal symptom that were recorded more than six weeks but less than three months apart. However, this excluded patients with a code for gastrointestinal symptoms and a free text note about the duration of ongoing symptoms. I, therefore, need to interpret the results with caution.

I included only the first recorded period of NICE eligible symptoms. Patients might have had additional eligible periods between 2013 and 2016. Therefore, the population of patients with eligible symptoms could be underestimated.

3.6.3 FINDINGS IN CONTEXT OF PUBLISHED EVIDENCE

3.6.3.1 IBD prevalence and incidence rates

Published figures on UK IBD incidence rates range from 21-37.5/100,000.^{23, 96-98, 113,}

¹¹⁴ Studies consistently report that prevalence is rising worldwide because of the low mortality associated with this chronic condition. UK prevalence estimates range from 328/100,000 in the 1990s²⁴ to 970/100,000 in 2017.⁹⁸

My estimates of incidence and prevalence of IBD in the UK are about 1.8 and 1.5 times higher than the most recently published estimates. Published studies are very heterogeneous, complicating comparison of reported rates across studies. Major variations that explain at least some of the differences include: 1) My study included adult patients only, while the majority of other studies covered a wider age range including children. This impacts the incidence and prevalence rates of IBD which has an onset that peaks in adulthood. 2) Improvements in diagnostic technology now enable the detection of milder cases.¹¹⁵ 3) Some smaller studies used GP records to identify cases with subsequent exclusion of unverified cases. Exclusions ranged from 8-26% of patients.^{23, 96, 113} This could have underestimated true IBD prevalence. 4) Studies used different definitions of disease. A number of studies did not include indeterminate IBD or microscopic IBD in their definition. A recent study reported the incidence and prevalence of ulcerative colitis and Crohn's disease in the THIN database.⁹⁸ The study only included Read codes for Crohn's disease and ulcerative colitis in the definition of disease. In contrast I used a very comprehensive and sensitive list of Read codes and drug codes (48 codes) for the identification of IBD, ulcerative colitis, Crohn's disease, indeterminate IBD and microscopic colitis. In addition, a further study using the THIN data used a similar list of Read codes to my study for the identification of IBD. However, they included non-specific IBD medications to identify IBD cases and only included patients with at least two subsequent IBD records or an IBD record and a recorded prescription of an IBD related drug.⁹⁷ According to my data, this approach may have missed at least 37.8% of cases. I was able to increase the sensitivity of my Read code list by using medications to identify additional IBD cases because I restricted inclusion of prescriptions to IBD specific medications. This is an advantage of my study over these two recent THIN studies.

Taken together, the evidence suggests that the IBD incidence and prevalence in the UK adult population may be higher than the latest published figures. IBD is a

heterogeneous group of disorders with Crohn's disease and ulcerative colitis considered as the two extremes of a spectrum of chronic gut disorders.¹¹⁶ My sensitive approach to identifying IBD cases may be more reflective of the true burden of disease in UK general practice.

3.6.3.2 Uptake of FC testing

My analysis showed a noticeable increase of FC test use by THIN practices from 2013 onwards, which coincided with the NICE guidance on faecal calprotectin.⁵⁹ However, an effect of NICE guidance on new practices starting FC testing was not seen. Overall, uptake was slow with 43% of THIN practices without an FC test record by 2016. The theory of diffusion of innovation explains that most innovations diffuse at a slow rate.¹¹⁷ This is because adopters, CCGs and GPs, weigh up the uncertainties of the consequences of the innovation, the FC test, against the advantages it can provide to solve a perceived need.¹¹⁷ Adoption is supported by providing information on the technology's consequences, advantages and disadvantages rather than just information on the technology itself.¹¹⁷ The NICE guidance does not provide this level of information. A consensus paper from the Faecal Calprotectin Working Group met some of the criteria.¹¹⁸ The consensus paper was produced with the aim to improve adoption and consensus on FC test use.¹¹⁸ It was based on experiences from 14 early adopter sites. Their algorithm is more specific than the NICE guidance, explicitly stating indications for testing and recommendations for urgent referral without testing for cases where acute IBD is suspected. It also includes a clear age range for testing from 18-60 years if cancer is not suspected. And it provides clear guidance on how to act on the FC test result according to a 100µg/g and 250µg/g FC threshold.

However, the document failed to address the question of test accuracy of FC testing in primary care, it did not explore why GPs choose not to use FC testing or why GPs refer test negative patients for further investigation. GPs appear to have different thresholds for testing and different reasons for testing according to the variation in tests ordered per 1,000 practice patients. However, as I do not know the optimal threshold for testing, I was unable to judge which number of tests per 1,000 patients would be the most appropriate. An understanding of who GPs test would be a step towards this and I explore this further in the next section and in Chapter 5.

3.6.3.3 Characterisation of FC test use

It is difficult to conclude from my THIN analysis who GPs test. The large proportion of FC tests taken in patients without eligible symptoms and without subsequent IBD diagnosis may indicate inexperience on the part of GPs in who to test. I could not determine reasons for testing in primary care because reporting of abdominal symptoms prior to FC testing was poor. Only 3.1% of patients with eligible symptoms received FC testing. Knowledge of presenting symptoms was more complete in published studies where indications for testing were actively sought.^{55, 64, 65} Published studies agreed that patients with symptoms of bloating or constipation were not frequently tested with FC and that the main indications for testing were abdominal pain and diarrhoea.^{55, 64, 65} Testing was sometimes indicated by unintentional weight loss and rectal bleeding.^{64, 65} I did not consider these symptoms because they were not indicators for FC testing according to the NICE guidance on FC testing. A prospective study design may be more suited to study reasons for testing.

The NICE evaluation of FC testing recommended FC testing for patients under 45 years of age. However, in my study about 20% of tests were carried out in patients aged 60 years and over. Furthermore, 27.8% of IBD diagnoses in FC tested patients were in patients 50 years and over. This goes in hand with the second reported peak of Crohn's disease onset between 50 and 60 years of age. My findings support the revised cut-off of 60 years for FC testing in primary care as suggested by the Faecal Calprotectin Working Group.¹¹⁸ NICE guidance may need revising.

3.6.3.4 Referral and colonoscopy following FC testing

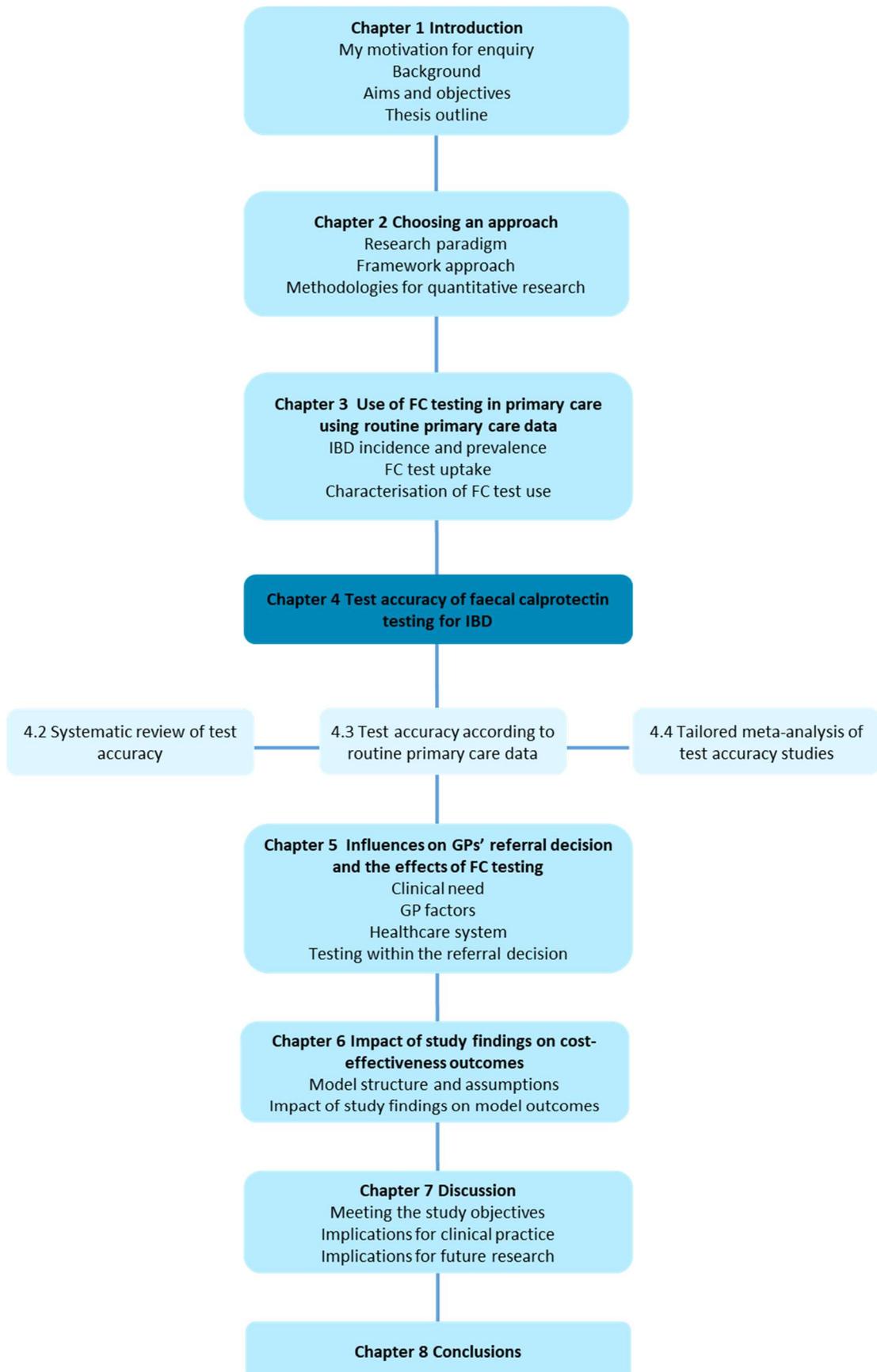
Numbers of FC tested patients referred and colonoscoped are in line with published figures from primary care studies. Published studies consistently showed that between 20% and 45% of patients with low FC levels were referred to gastroenterology.^{51, 55, 63-65} The THIN estimate was in line with the upper value of referred FC negative patients. This raises concerns about the impact of FC testing on colonoscopy rates as a considerable proportion of referred patients with negative FC levels were further investigated. Inappropriate use of FC testing as a screening test rather than to rule out IBD has been proposed as an explanation.¹¹⁹ However, the number of investigations of referred FC negatives varied greatly among my own (13.6%) and the three published studies (19%,⁶³ 46%⁵¹ and 71%⁵⁵). This suggests that the number of potentially unnecessary colonoscopies may be at least as

dependent on secondary care decision-making as on the availability of FC testing in primary care. None of the studies can explain the reasons for referring patients with low FC levels in whom IBD is unlikely. This suggests that GPs have other reasons for referral, which FC testing may not address.

3.7 Summarising the use of FC testing in primary care

In Chapter 3 I focused on the use of FC testing in primary care to characterise FC testing in light of NICE guidance. I showed that FC uptake was slow, that IBD prevalence and incidence were higher than reported in published studies, that reasons for testing were inadequately recorded in THIN and that GPs referred a proportion of patients with FC negative test results that cannot be explained by test results alone. Non-compliance with NICE guidance may suggest that recommendations lack applicability to the primary care context.

The THIN data were insufficient in explaining who GPs test and why, or what the reasons were for referral of FC negative patients. I explore these questions using qualitative interviews in Chapter 5. First, however, I address the issue of test accuracy of FC testing in Chapter 4. I investigate the test's accuracy in primary care compared to secondary care to be able to judge what weight GPs can place on the FC test result in the decision whether to refer or manage a patient with abdominal symptoms.



Chapter 4 Test accuracy of faecal calprotectin testing for IBD

4.1 Chapter overview

In Chapter 3 I investigated and characterised the use of FC testing in primary care. I showed that uptake and use was inconsistent. I suggested that this may be due to a lack of experience of GPs in FC testing. This may have resulted in reduced confidence in test results, especially when guidance is based on secondary care test accuracy studies. This chapter explores three approaches to establishing the test accuracy of FC testing for the detection of IBD in primary care.

Firstly, I present the findings of my systematic review of published test accuracy studies, from both secondary and primary care settings. I compare the evidence from the two different settings comprising different patient populations with different spectra of disease and different co-morbidities. The meta-analysis was exploratory, mainly due to heterogeneity from the different clinical questions addressed and the different testing assays employed by the different studies. The meta-analysis aimed to establish the range of possible test accuracy values rather than produce one overall test accuracy estimate.

Secondly, I report the results of my primary care test accuracy study. I explore the use of routine electronic health records to address the question of test accuracy of FC testing in primary care. This section looks first at the data management strategies needed to formulate assumptions and inclusion criteria for the inclusion of cases of FC testing into the test accuracy study, illustrating the shortcomings and problems when dealing with routine data. After that, I present the findings of the test accuracy study and sensitivity analyses, which test some of my main assumptions.

Thirdly, I describe my tailored meta-analysis as an alternative method to estimate the test accuracy of FC testing for IBD for the primary care setting. The tailored meta-analysis approach combines information from published evidence on test accuracy (the systematic review) with routine data from the setting of interest (the test accuracy study) and aims to produce more applicable estimates of sensitivity and specificity by tailoring the inclusion of studies to those deemed most plausible to the primary care setting. By comparing the findings to the results of the THIN test accuracy study I attempt to validate the tailored meta-analysis method for the first time.

4.2 Systematic review and exploratory meta-analysis of diagnostic test accuracy studies of FC testing

This section is published as:

Freeman K, Willis BH, Fraser H, Taylor-Phillips, S and Clarke A. *Faecal calprotectin to detect inflammatory bowel disease: a systematic review and exploratory meta-analysis of test accuracy*. BMJ Open 2019; 9:e027428.

4.2.1 INTRODUCTION

Slow uptake of FC testing into general practice and the referral of test negative patients into specialist care, which I reported in Chapter 3, might be due to a lack of confidence in the test accuracy of FC testing, especially without compelling evidence from the primary care setting. The FC test is recommended by gastroenterological societies across the globe for its usefulness in the diagnosis of IBD.¹²⁰⁻¹²³ However, there is no clear guidance on the settings in which it should be considered appropriate. FC testing is approved for the differential diagnosis of IBD and IBS in adult primary care patients in the UK. NICE recommends FC testing when referral to secondary care is being considered and cancer is not suspected (DG11).⁵⁹ This guidance was based on a high quality systematic review and health economic evaluation with the majority of the evidence coming from secondary care.⁵⁸ Insights into how the FC test may work in primary care came from two pilot studies funded by the NHS Technology Adoption Centre.⁶ The pilot studies involved two areas in the north of England in which GP practices implemented FC testing into their routine clinical practice for patients with suspected IBS and IBD (123 and 77 patients respectively). The working diagnosis and referral decision were not influenced by the FC test. The FC test result was compared with the consultant diagnosis. However, the studies did not qualify for inclusion into the review because 1) the consultant diagnosis included information of the FC test and 2) not all patients received colonoscopy. The findings from these pilot studies, therefore, concerned mainly the referral pattern of patients with abdominal symptoms rather than the test accuracy. Test accuracy measures of the review were informed by studies from secondary care with a different spectrum of disease. Identifying the most appropriate patient population for testing was problematic because of the different populations tested in the studies. Waugh et al. (2013) classified the populations into four clinical decision questions: IBD versus IBS, organic disease versus IBS, inflammatory disease versus non-inflammatory diseases, and organic versus non-

organic disease.⁵⁸ They stated that the differential diagnosis in the adult primary care population would be mainly IBD in a background of IBS. The performance of the FC test was most promising in differentiating IBD from IBS with a pooled sensitivity and specificity from five studies of 0.93 (95% CI 0.83 to 0.97) and 0.94 (95% CI 0.73 to 0.99), respectively.⁵⁸ However, this patient population was not representative of primary care because it was based on the exclusion of other conditions. Our knowledge of how the test performs in primary care with a more heterogeneous population including patients with mild symptoms is limited, and a systematic approach to assessing the applicability of the available evidence specifically for this setting is needed.

4.2.2 OBJECTIVES

- To assess the test accuracy of FC testing to detect IBD in adult patients with chronic abdominal symptoms in a primary care pathway.
- To explore the differences in FC test performance between primary and secondary care.

4.2.3 METHODS

4.2.3.1 Review methods

I followed the protocol by Waugh et al. (2013)⁵⁸ registered as PROSPERO CRD 42012003287 except that I excluded studies on children which were not within the scope of the NICE guidance.⁵⁹ I considered all studies included by Waugh et al. (2013)⁵⁸ for inclusion and searched MEDLINE, EMBASE, the Cochrane Library and Web of Science from 01/09/2012 to 31/05/2017 (Appendix B1). Eligible studies from auto-alerts were included up until 31/01/2018 and I checked all reference lists of included studies.

Studies were assessed for eligibility independently and in duplicate, and any disagreements were resolved by consensus. I documented records rejected at full text stage and reasons for exclusion.

I included studies which measured FC levels in stool samples to detect IBD in adult patients (with ≥80% of study population 18-60 years) with chronic (at least 6-8 weeks) abdominal symptoms not yet diagnosed in primary or secondary care. The reference standard to verify FC test outcomes was colonoscopy with histology,

other imaging technologies and follow-up. The main outcomes of interest were sensitivity and specificity.

I extracted data on pre-specified data extraction sheets, which were checked by a second reviewer. For studies reporting sensitivity, specificity, positive and negative predictive values and total number of included patients, 2x2 tables of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) were calculated. For studies reporting test accuracy at a lower and an upper cut-off, 2x2 data tables were extracted for the lower and the upper cut-off. I contacted authors of relevant studies with missing 2x2 data to request data.

In contrast to the previous review, firstly, I disregarded repeat testing in one study¹²⁴ in the formulation of the 2x2 table. I believe this is more appropriate and relevant for the performance of the test in clinical practice as a decision needs to be made on the first test result. Secondly, I excluded data investigating FC testing in small bowel disease in comparison to radiology in one study.¹²⁵ This study included only endoscopy negative patients and, therefore, addressed a different question in a different patient population using a different reference standard to the review question. Finally, I did not simply adopt the review's categorisation of studies into the same clinical decision questions. I assessed all included studies and identified six different clinical questions which I subsequently categorised into three groups.

1) IBD versus IBS

IBD versus IBS and IBD versus functional disease were assumed to be equivalent because I expected the vast majority of functional disease in this context to be IBS.

2) IBD versus non-IBD

I assumed that inflammatory versus non-inflammatory disease is similar to the category of IBD versus non-IBD accepting a few misclassifications according to the different definition of the diseased and non-diseased groups. (This can be demonstrated using the example of diverticulitis. Diverticulitis is a non-IBD inflammatory condition. In the differentiation of IBD versus non-IBD diverticulitis would be categorised as non-IBD and a positive test would contribute to the false positive results. In the differentiation of inflammatory versus non-inflammatory disease diverticulitis would be classed as the target condition and a positive test would contribute to the true positives.) This group included publications of clinical experiences using

routine data to evaluate the value of FC testing for IBD. These studies were of particular interest to me because I had planned a similar approach for my THIN test accuracy study in section 4.3.

3) Organic versus non-organic disease

Organic disease (versus non-organic or versus 'other') was too broad for the review question as a number of conditions listed by the studies were not indications for FC testing and the outcomes were less likely to be test specific. This clinical question was secondary to the other two questions.

Quality was assessed independently and in duplicate by myself and a second reviewer using the tailored Quality Assessment of Diagnostic Accuracy Studies-2 criteria (QUADAS-2)¹²⁶ which included definitions for the signalling questions to match the review question.

4.2.3.2 Analysis

I considered studies of FC testing under three main clinical questions: 1) IBD versus IBS, 2) IBD versus non-IBD conditions and 3) organic versus non-organic conditions. Different FC test assays were handled independently.

I used Review Manager 5.3 (The Cochrane Collaboration, The Nordic Cochrane Centre, Copenhagen, Denmark) to produce paired forest plots of sensitivity and specificity and performed analyses in R version 3.4.1 (Vienna, Austria).¹⁰⁴ I undertook meta-analyses using a bivariate random effects model¹²⁷ using the function *glmer* in the package *lme4*¹²⁸ with a minimum of five studies required for meta-analysis. I considered different assays and clinical questions separately and produced no overall summary estimates of sensitivity and specificity due to heterogeneity. I explored test accuracy at the commonly used thresholds, the value at which the test is considered positive, of 50 and 100 µg faecal calprotectin per g stool sample (µg/g). To produce receiver operating characteristic (ROC) plots I entered outputs from meta-analyses into Review Manager.

Heterogeneity was explored by meta-regression analyses with assay type and clinical question added as covariates in turn. I assumed equal variances (Appendix B2) because of small numbers of studies per subgroup. Additional models assuming unequal variances did not converge.

Each meta-analysis considered only one outcome per study. However, a number of studies reported 2x2 data for multiple test assays and clinical questions. I, therefore, undertook an exploratory sensitivity analysis. First, I produced a list of all possible combinations of test assay and clinical question for each included study at a common threshold (some studies reported outcomes for one test assay and clinical question contributing one possible combination while others reported results for up to three test assays and two clinical questions resulting in 6 possible combinations). I then ran a random sample of 25,000 meta-analyses out of a possible 10 million picking one outcome per study at random in each round of meta-analysis and considering all combinations of test assays and clinical questions reported in the studies. Pairs of sensitivity and specificity from these meta-analyses were plotted in a ROC plot and a 2D density plot to take account of the different definitions of disease and test assays and to allow visualisation of the variability in the data. The R code for this analysis is available in Appendix B3.

4.2.4 RESULTS

Figure 4.1 provides the PRISMA¹²⁹ flow diagram of study selection. Appendix B4 lists the records rejected at full text stage and reasons for exclusion. Out of 2,168 unique records 38 studies were eligible for inclusion. The summary Table 4.1 reveals extensive heterogeneity in all major aspects of test accuracy studies. Study characteristics by study can be found in Appendix B5.

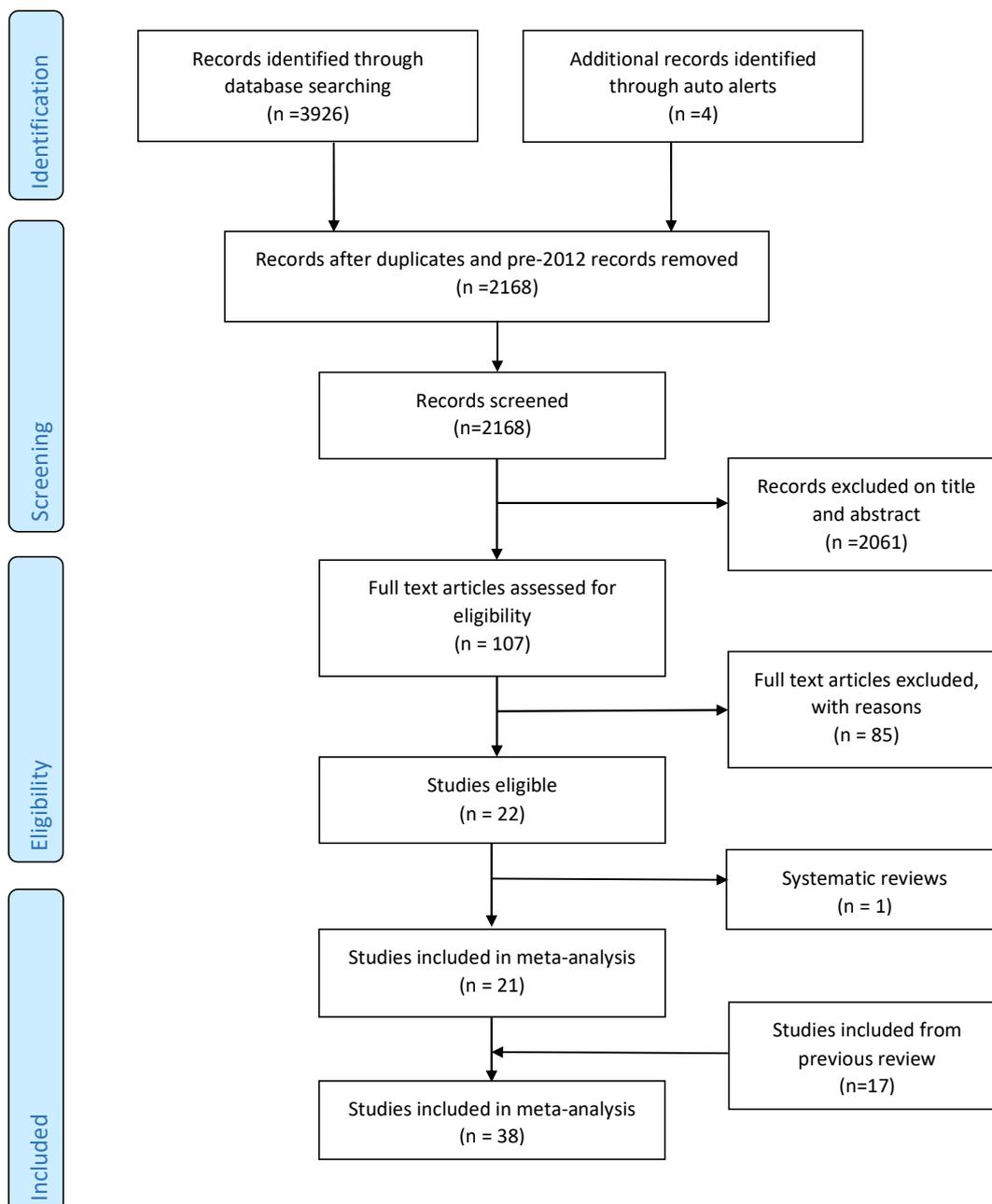


Figure 4.1 Flow diagram of study inclusion

Table 4.1 Summary of study characteristics of included studies addressing the test accuracy question of FC testing in the detection of IBD

Study characteristic	Outcome
Study characteristics:	
Publication type (studies)	
Full text	32
Abstract	6
Year of publication (range)	2000-2018
Population size (range)	31-1,031
Geographical region (studies):	
UK	14
Rest of Europe	16

Asia	2
Middle East	2
North Africa	1
Russia	1
Canada	1
USA	1
Patient characteristics:	
Age (range)	14-97 years
IBD prevalence (range)	2.1% to 76%
FC assay type (studies)*:	
Laboratory immunoassays	
ELISA	32
FEIA	5
CLIA	3
PETIA	2
POCT	10
Setting (studies)	
Primary care	5
Secondary care	29
Out- and inpatients	12
Referred patients	17
Mix	2
Unclear	2
Target condition (studies)*:	
IBD	23
Inflammatory disease	5
Organic disease	19
Non-target condition (studies)*:	
IBS	12
Functional disease	2
Non-IBD	12
Non-organic disease	16
Non-inflammatory disease	5
Other	3
FC test data collection (studies):	
Prospectively for patients with eligible symptoms in primary care	1
Prospectively for patients at time of referral in primary care	1
Retrospectively of routine FC tests in primary or secondary care	10
Prospectively prior to a planned colonoscopy in secondary care	23
Prospectively during the assessment for the need of colonoscopy in secondary care	1
Unclear	2
Reference standard (studies):	
Colonoscopy with biopsy	13
Colonoscopy +/- biopsy	7
Endoscopy + other imaging tests	8
Endoscopy + follow-up	3
Endoscopy + other imaging tests +follow-up	5
Unclear	2

*Some studies evaluated multiple tests/clinical questions

CLIA chemiluminescent immunoassay, ELISA enzyme-linked immunosorbent assay (=EIA enzyme immunoassay), FEIA fluorescence enzyme immunoassay, IBD inflammatory bowel disease, PETIA particle enhanced turbimetric immunoassay, POCT point of care test

Figure 4.2 summarises the assessment of risk of bias and applicability (see Appendix B6 for the assessment by study). Overall, the risk of bias was high or uncertain across all four domains in at least 50% of studies. Concerns about the applicability of the patient population to a primary care setting were high in about 75% of studies.

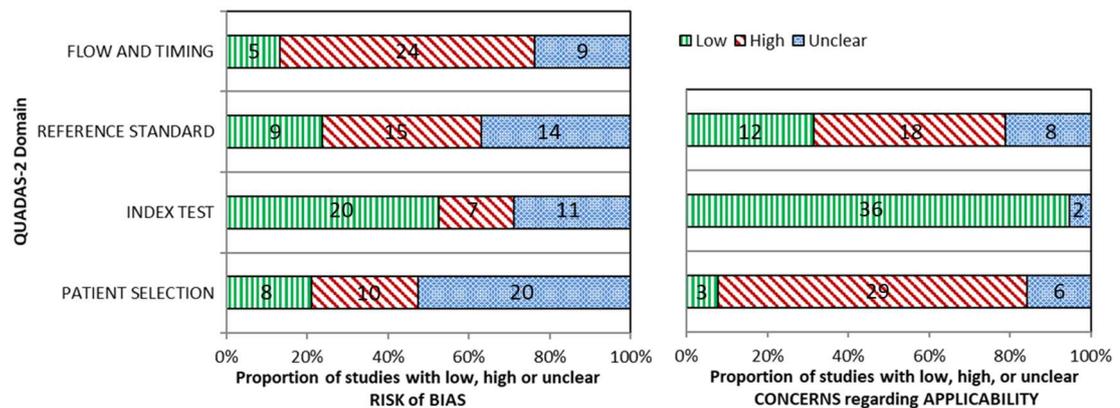


Figure 4.2 Overview of risk of bias and applicability concern of included studies according to QUADAS-2¹²⁶ (numbers represent numbers of studies in each category)

The Forest plots in Appendix B7 describe the complete evidence on test accuracy from the 38 included studies reporting all the thresholds, assays and clinical questions explored. Overall, sensitivity decreased and specificity increased as the threshold increased. FC testing appears to be more accurate in the detection of the more precise clinical construct of IBD than in the detection of organic disease as a whole.

Two studies compared five and six different test assays at a common threshold of 50µg/g.^{130, 131} This allowed for a test comparison in the same study population under similar study conditions. It appeared here that the specificity varies to a greater extent than the sensitivity (Figure 4.3). Both studies reported medium to strong correlations but low agreement between assays, meaning that analysing the same sample with different FC test assays will result in different values for FC.

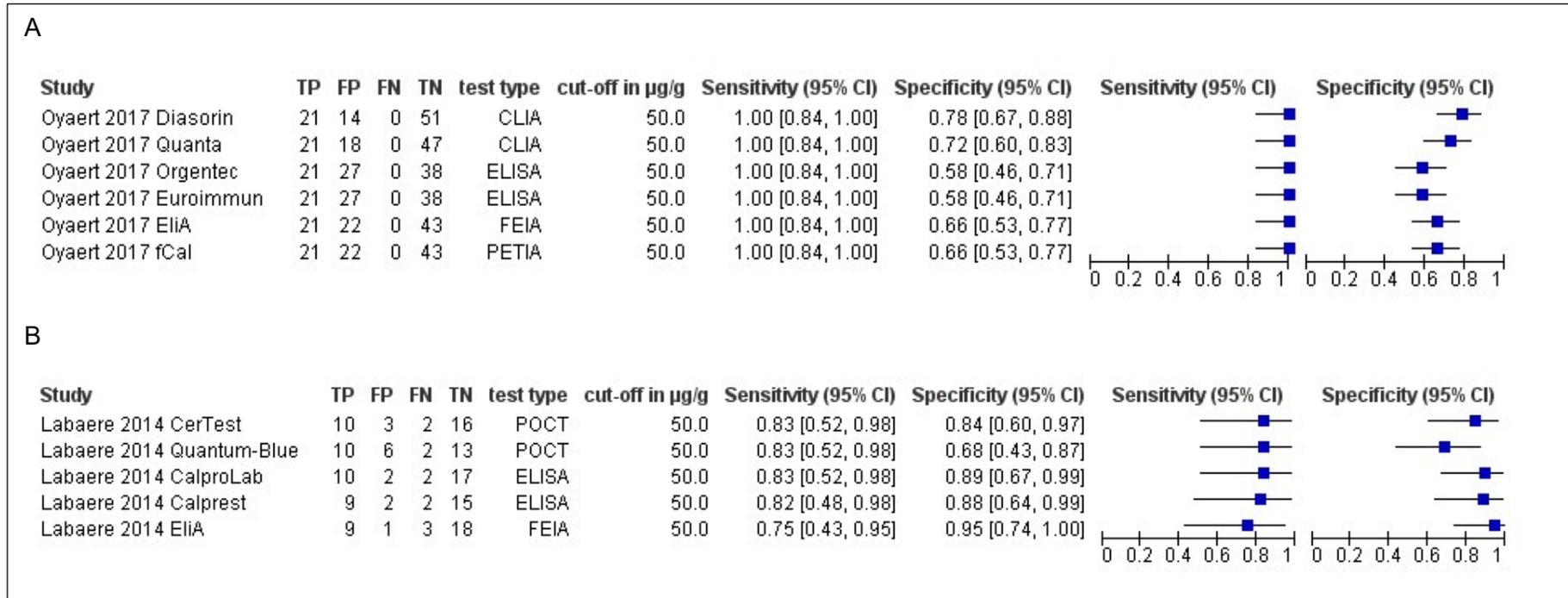


Figure 4.3 Forest plot of A) Oyaert 2017¹³⁰ comparing 6 FC tests and B) Labaere 2014¹³¹ comparing five FC tests at cut-off 50 $\mu\text{g/g}$
 CI confidence interval, CLIA chemiluminescent immunoassay, ELISA enzyme-linked immunosorbent assay, FEIA fluorescence enzyme immunoassay, FN false negative, FP false positive, PETIA particle enhanced turbimetric immunoassay, POCT point of care test, TN true negative, TP true positive

4.2.4.1 Meta-analysis at 50µg/g threshold

28/38 studies reported test accuracy at the common 50µg/g cut-off and could be considered for meta-analysis. Meta-analyses were undertaken separately for test assays and for clinical questions to explore heterogeneity, and to allow for multiple outcomes from a number of studies.

All 18 assays were considered in the meta-analysis, investigating the effect of assay type on summary estimates of test accuracy. (This was based on the assumption that the clinical question is generic and one outcome was picked per study for analysis). Four assays (PhiCal, EK-CAL, Quantum-Blue, and EliA) with five or more studies each could be considered in the comparison (Appendix B8 for 2x2 data). Nineteen studies contributed data.^{33, 40, 50-52, 106, 124, 130-141} Five studies contributed data to two different assays.^{40, 52, 131, 135, 140} Figure 4.4 depicts the pairs of sensitivity and false positive rates (1-specificity) of contributing studies in the ROC space, with summary estimates by assay type. At the global 50µg/g threshold, test performance appeared to vary slightly across assays (Table 4.2). Quantum-Blue had the highest summary estimate for sensitivity with 0.94 (95% CI 0.75 to 0.99) but also the lowest specificity (0.67, 95% CI 0.56 to 0.76). The greatest difference in sensitivity (9%) was between EK-CAL ELISA and Quantum-Blue point of care test (POCT) while the greatest difference in specificity (21%) was between PhiCal ELISA and Quantum-Blue POCT, again suggesting greater variation in specificity than in sensitivity across assays. Whether these differences are due to differing performance of test assays or due to methodological issues in study designs is difficult to ascertain. The differences are large enough to suggest that test assays should not be treated as equivalent.

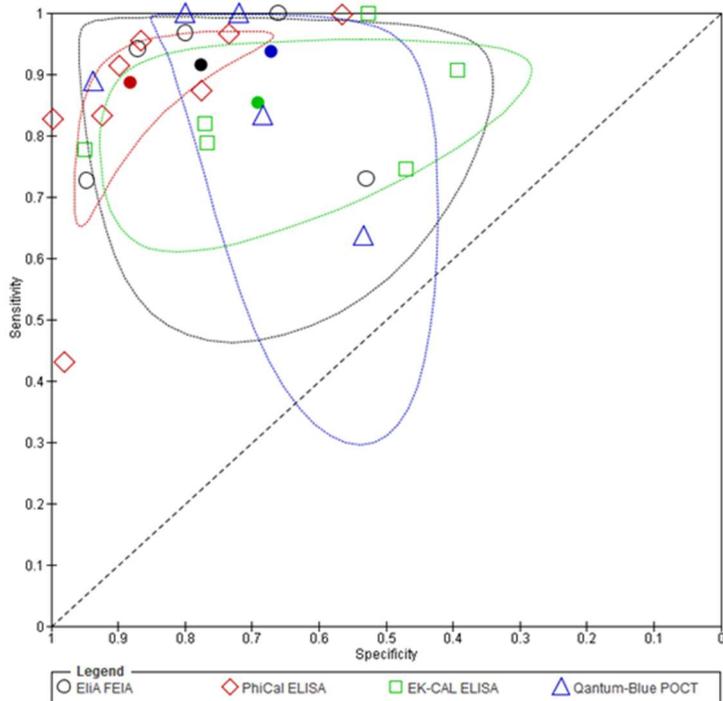


Figure 4.4 Receiver operating characteristic (ROC) plot of sensitivity and specificity by test (filled shapes present summary estimates with 95% confidence region)

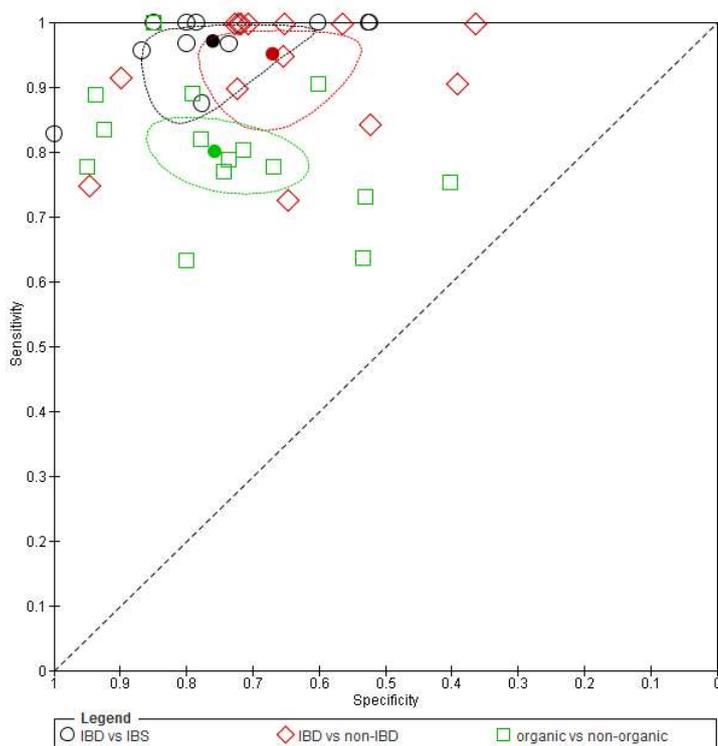


Figure 4.5 Receiver operating characteristic (ROC) plot of sensitivity and specificity by clinical question (filled shapes present summary estimates with 95% confidence region)

Table 4.2 Comparison of results of separate meta-analyses and meta-regression with test type and clinical question as covariates at two thresholds (numbers in brackets are numbers of studies meta-analysed)

Test comparison at 50µg/g threshold								
Method	PhiCal (n=8)		EK-CAL (n=6)		Quantum-Blue (n=5)		ELiA (n=5)	
	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Separate MA for each test	0.89 (0.76; 0.95)	0.88 (0.77; 0.94)	0.85 (0.75; 0.92)	0.69 (0.47; 0.85)	0.94 (0.75; 0.90)	0.67 (0.56; 0.76)	0.92 (0.78; 0.97)	0.78 (0.60; 0.89)
Meta-regression (test) with equal variances	0.89 (0.77; 0.95)	0.89 (0.79; 0.95)	0.87 (0.71; 0.95)	0.69 (0.49; 0.84)	0.92 (0.78; 0.98)	0.73 (0.50; 0.88)	0.92 (0.77; 0.97)	0.78 (0.57; 0.91)
Comparison of clinical questions at 50µg/g threshold								
Method	IBD vs IBS (n=11)		IBD vs non-IBD (n=14)		Organic vs non-organic (n=15)			
	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Separate MA for each clinical question	0.97 (0.91; 0.99)	0.76 (0.66; 0.84)	0.95 (0.88; 0.98)	0.67 (0.58; 0.75)	0.80 (0.76; 0.84)	0.76 (0.66; 0.83)		
Meta-regression (clinical question) with equal variances	0.96 (0.92; 0.98)	0.77 (0.66; 0.85)	0.94 (0.90; 0.97)	0.67 (0.57; 0.76)	0.81 (0.74; 0.86)	0.75 (0.67; 0.82)		
Comparison of clinical questions at 100µg/g threshold								
Method	IBD vs IBS (5)		IBD vs non-IBD (5)		Organic vs non-organic (7)			
	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Separate MA for each clinical question	0.92 (0.85; 0.96)	0.86 (0.82; 0.89)	0.72 (0.63; 0.80)	0.82 (0.78; 0.86)	0.67 (0.56; 0.76)	0.87 (0.84; 0.90)		
Meta-regression (clinical question) with equal variances	0.92 (0.85; 0.96)	0.86 (0.81; 0.89)	0.73 (0.60; 0.82)	0.82 (0.77; 0.86)	0.67 (0.58; 0.75)	0.87 (0.84; 0.89)		

CI confidence interval, IBD inflammatory bowel disease, IBS irritable bowel syndrome, MA meta-analysis, vs versus

All reported 2x2 data (Appendix B9) for the clinical questions IBD versus IBS, IBD versus non-IBD, and organic versus non-organic disease were considered in the meta-analysis investigating the effects of the different clinical questions on summary estimates of test accuracy under the assumption that the test is generic. Twenty eight studies were included at the common threshold of 50µg/g.^{33, 40, 50-52, 54, 61-63, 106, 124, 130-146} Ten studies contributed information to more than one category.^{50, 54, 62, 63, 134, 136-138, 143, 144} The ROC plot in Figure 4.5 summarises the summary estimates for the comparison of the three clinical questions and the pairs of sensitivity and false positive rates for the contributing studies. Table 4.2 shows that the point estimate of specificity for IBD versus non-IBD (0.67, 95% CI 0.58 to 0.75) is lower than for IBD versus IBS (0.76, 95% CI 0.66 to 0.84). Acknowledging that the confidence intervals overlap, this suggests that broadening the definition of the non-target condition may produce more false positives as other non-IBD intestinal conditions are included. On the other hand, sensitivity for organic disease (0.80, 95% CI 0.76 to 0.84) is lower than for IBD (0.97, 95% CI 0.91 to 0.99). This suggests that some organic disease will be missed with FC testing as organic conditions, including adenomas and diverticulosis, are not typically associated with inflammation. Considering different definitions of the clinical questions results in greater variation in sensitivity than specificity meaning that widening the target condition had a greater impact on test accuracy than changing the non-target condition (IBS versus non-IBD versus non-organic disease).

4.2.4.2 Meta-analysis at 100µg/g threshold

Eleven studies reported test accuracy at a threshold of 100µg/g (Appendix B10 for 2x2 data).^{50, 51, 54, 55, 62, 63, 134, 143, 144, 147, 148} Of these, five contributed to more than one category of clinical questions.^{50, 54, 62, 134, 144} As expected, raising the threshold to 100µg/g reduced sensitivity and increased specificity for all clinical questions (Table 4.2). Insufficient numbers of studies at the 100µg/g threshold were available to explore assay type.

4.2.4.3 Meta-regression and exploratory sensitivity analysis

Results from separate meta-analyses and from meta-regression analyses with equal variances showed similar results (Table 4.2).

Ten studies reported results for multiple clinical questions^{50, 54, 62, 63, 134, 136-138, 143, 144} and eight studies compared multiple tests.^{40, 52, 131, 133, 135, 136, 138, 140} My approach of

randomly selecting one test or one clinical question per study for meta-analysis uses only one set of evidence per study. Choosing a different set might have resulted in different outcomes and conclusions as the meta-analyses suggested that it might not be appropriate to consider tests and questions alike. In an attempt to capture this variation, Figure 4.6 displays results of 25,000 meta-analyses of 28 studies (Appendix B11 for 2x2 data) picking one outcome per study at random for each round of meta-analysis. The results show overall high sensitivity and specificity irrespective of test assay and clinical question (Panel A) with slightly greater variation in sensitivity than specificity (Panel B). Median sensitivity and specificity of the 25,000 analyses and ranges were 0.9 (min 0.85, max 0.94) and 0.76 (min 0.73, max 0.79), respectively.

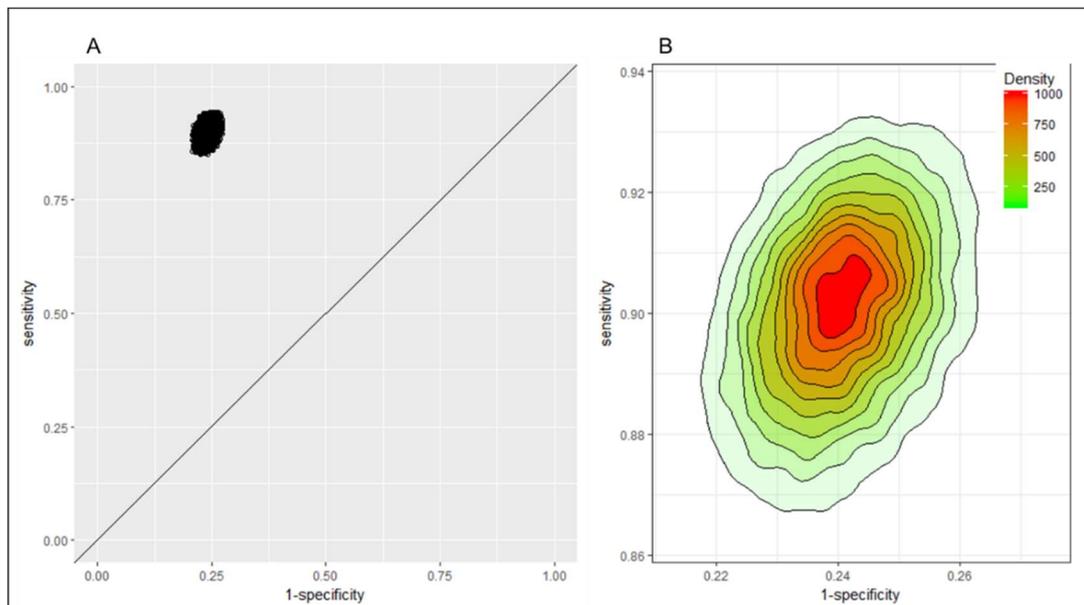


Figure 4.6 Pairs of sensitivity and specificity of 25,000 meta-analyses of 28 studies picking one outcome per study at random for each round of meta-analysis at a threshold of 50 μ g/g. A) Receiver operating characteristic plot, B) 2D density plot scaled to illustrate spread and density (where the highest density represents a probability of 1,000/25,000)

4.2.4.4 Primary versus secondary care

Five of the eligible studies were from primary care settings.^{51, 55, 61-63} The studies were too heterogeneous for meta-analysis. In an attempt to compare test performance of FC testing in primary care with secondary care only one study qualified.⁶³ The study evaluated the routine use of FC testing in primary care and reported lower sensitivity of 0.72 but comparable specificity of 0.65 for differentiating IBD from non-IBD at a threshold of 50 μ g/g when compared to pooled estimates across different test assays including all settings (sensitivity 0.95, 95% CI 0.88 to

0.98 and specificity 0.67, 95% CI 0.58 to 0.75). However, indication for testing and the place of the FC test in the patient pathway were unclear in this study, and the study may suffer from differential verification bias due to concerns over the reference standard. The authors explained the low sensitivity with the fact that general practitioners may have referred patients with high suspicion of IBD without testing. Another reasonable explanation could be the greater number of apparently milder IBD cases in primary care in which FC testing is falsely negative, reducing sensitivity in this setting, and demonstrating the impact of the different spectrum of disease on test accuracy estimates.⁶⁷

4.2.5 DISCUSSION

4.2.5.1 Study results

None of the 38 included studies sufficiently addressed the review question of test accuracy of FC testing in a primary care pathway. The studies recruited patients with a different spectrum of disease from that anticipated for primary care, focussed on disease groups which were broader than the intended IBD group, or did not verify all patients with the preferred reference standard. Furthermore, the evidence was highly heterogeneous. Studies varied in thresholds and assays used, as well as in target conditions and settings. Eighteen different test assays were studied. Although similar, these could not be considered the same for the purposes of meta-analysis. Clinical questions were analysed as three distinct categories which were related but which could not be pooled for meta-analysis. A number of studies compared multiple tests and/or explored different clinical questions. Picking one test and one clinical question per study for meta-analysis was unjustifiable. I favoured an innovative approach which allowed me to explore the breadth of evidence and showed that irrespective of clinical question and test assay, test accuracy of FC testing was high. However, the variation in specificity could translate into considerable uncertainty of false positives when scaled up to population testing, while the variation in sensitivity reflects the limited value of FC testing for other organic conditions.

4.2.5.2 Study limitations

The review has a number of limitations. Due to heterogeneity I decided not to pool across tests and clinical questions. Instead I conducted exploratory analyses which demonstrated that the test assay and clinical question may affect the summary sensitivity and specificity by as much as 9%. I was also unable to compare primary

and secondary care studies due to the small number of heterogeneous studies in primary care. The meta-regression analyses assumed equal variances between categories of tests and clinical questions, and this was supported by a statistical comparison. However, the power of this test is dependent on the number of studies per sub-group and it is possible that this assumption could be challenged with a larger data set. Finally, the categorisation into different clinical questions was subjective because the disease categories are ill defined and studies' definitions of conditions and groups of conditions varied.

4.2.5.3 Findings in the context of published studies

This review is broadly in line with the approaches and interpretation of the evidence of the previous review.⁵⁸ However, Waugh et al. (2013)⁵⁸ reported pooled estimates of sensitivity (0.93, 95% CI 0.83 to 0.97) and specificity (0.94, 95% CI 0.73 to 0.99) across tests for the differential diagnosis of IBD versus IBS for five secondary care studies. This was the basis for national decisions to introduce FC testing in primary care. When compared to my equivalent analysis of 11 studies I found comparable sensitivity (0.97, 96% CI 0.91 to 0.99) but considerably lower specificity (0.76, 95% CI 0.66 to 0.84) suggesting that false positives might be more of a concern than previously concluded. Thus, the predicted reduction in colonoscopies and subsequent cost savings may not be realised as a result of introducing FC testing into the primary care pathway in the UK.⁵⁸ Furthermore, this review found enough disagreement between tests to caution against treating tests as equivalent. Issues in homogenisation, dilutions and extraction prior to analysis as well as lack of standardisation of different assays contribute to these differences. Recommended cut-off values would have to be determined locally until these issues are resolved.¹³⁰

Only five studies recruited patients from primary care populations and analysed the general practitioners' decisions on whether to test and whether to refer.^{51, 55, 61-63} I found lower sensitivity for detection of disease in these studies compared to meta-analyses of all studies. None of the included primary care studies assessed FC testing for the differential diagnosis of IBD versus IBS. A more recent study, however, attempted the comparison by excluding non-IBD organic disease from the analysis.⁶⁴ The comparison of IBD versus non-IBD might reflect the clinical situation for FC testing more accurately, but produces more false positives as FC levels might be raised in diverticular disease, coeliac disease, rectal adenocarcinoma, non-specific inflammation, and others.^{51, 55} The meaning of the additional false positives due to non-IBD inflammatory conditions is debatable in the context of

clinical practice where incidental findings of true disease would not be classed as a false positive test outcome but would prompt further investigations, hence the effort by Walker et al. (2018) to include all organic intestinal disease in the target condition in a second analysis.⁶⁴

NICE recently endorsed a cut-off of 100µg/g for use in primary care in England and Wales, based on a primary care study which showed a 43% reduction in FC test positives compared with a threshold of 50µg/g.⁶³ I investigated test accuracy at this cut-off including all settings and demonstrated an increase in specificity by at least 10%, while the magnitude of the resulting decrease in sensitivity was more uncertain. A more recent study from primary care advocated the new threshold for referral decisions from primary care.⁶⁴ However, this study did not address the clinical question of FC testing distinguishing between IBD and non-IBD but explored the potential of FC testing to detect organic intestinal disease more general. More evidence on the test accuracy of FC testing in primary care to detect IBD using the revised cut-off is needed before any recommendations for use in primary care can be made.

4.3 Test accuracy of FC testing in primary care according to THIN

4.3.1 INTRODUCTION

Despite the publication of primary care studies since the review by Waugh et al. (2013),⁵⁸ which I described in my systematic review (section 4.2), we still lack compelling evidence on the test accuracy of FC testing in primary care for the detection of IBD. This is because the primary care evidence is based on small, heterogeneous studies with differences in clinical questions, FC thresholds, and definitions of disease. They report a range of sensitivities and specificities with wide confidence intervals and leave uncertainty over which threshold should be adopted in primary care. This is a concern given that FC testing is recommended for primary care by national guidance in the UK.⁵⁹ The decision to introduce FC testing in primary care might have been based on overoptimistic assessments of test accuracy in this setting. Evidence for the test accuracy of FC testing in primary care for different FC thresholds is needed before any recommendations for use in primary care can be made. The ability of FC to distinguish between IBD and IBS is known from test accuracy studies.¹²⁵ A more pragmatic approach is needed to support decisions on the use of FC testing in a specific setting, such as primary care. The focus should be on assessing the performance of FC testing in a large, multicentre study population that is representative of the setting and clinical question of interest.⁸⁴ This may require a more pragmatic approach to disease verification. A common reference standard is the follow-up of test negatives for evidence of a final diagnosis.¹⁴⁹ Such data are readily available in large routine GP databases such as THIN.

Routine electronic patient records are particularly useful for studying what is happening in the real world. Other advantages, relevant to test evaluation, include the availability of information related to the complete patient care pathway. This enables the evaluation of the added value of a diagnostic test in the diagnostic pathway.¹⁴⁹ Furthermore, the large numbers of patients enable studies of rare conditions.¹⁵⁰ The usefulness of routine data for test accuracy studies, however, has been questioned due to their inherent biases.¹⁵¹⁻¹⁵⁵ Reasons for biases include difficulties in the selection of the appropriate patient population based on symptoms, partial or differential verification, lack of blinding, missing data, missing variables (specifically reason for testing) and having to rely on clinical coding.^{149, 150} However, the increasing availability of routine data has shifted the focus from identifying biases to the question of how to deal with the problems. Solutions include the use of

symptom and problem oriented databases (for instance from general practice), pragmatic approaches to verification such as follow-up and consensus, multiple imputation of missing data, and judgement of the impact of bias.^{149, 150}

Previous studies which used routine data for the test accuracy of FC for IBD were based on data from a single medical centre,^{50, 63} a single laboratory⁶⁵ or an American health plans database.¹⁵⁶ In fact, I could only identify one former test accuracy study using a UK GP database¹⁰⁹ (and two studies on disease risk prediction algorithms^{149, 157}). Miller et al. (2015) investigated the diagnostic accuracy of rheumatoid factor as a test for rheumatoid arthritis using the CPRD.¹⁰⁹ The study's biggest risk of bias was the preferential recording of positive test results over negative test results, which improved throughout the study as reporting moved from manual to automatic.^{109, 110} A second source of bias was the potential misclassification of cases by using an algorithm for case identification with a sensitivity and specificity of about 85%. An advantage was that, in addition to reporting estimates of test accuracy, the study could investigate the use and impact of testing because of the information available in routine data.

I reported the test use and role of FC testing in patient management using the THIN data in Chapter 3. The aim of this study was to use the THIN data to explore the feasibility of using routine data to estimate test accuracy measures for FC testing for IBD in primary care.

4.3.2 OBJECTIVES

- To estimate the test accuracy of FC testing for IBD using THIN data
- To explore the impact of test results on referral
- To compare THIN estimates with published evidence from primary and secondary care

4.3.3 METHODS

The THIN study design and the methods for creating the THIN datasets were described in section 3.4, leading to the development of a dataset of 6,965,853 adult patients. For the test accuracy study I identified all first time FC tests in the dataset. I interrogated the records before considering them for inclusion into the study. Therefore, I first report the methods to decide which FC tests to include into the test

accuracy study. This is followed by a detailed description of the methods of the THIN test accuracy study and justifications for sensitivity analyses.

4.3.3.1 Methods of data management to set criteria for inclusion of patients into the test accuracy study

I used descriptive statistics such as frequencies, means, ranges and histograms to explore FC tests that were available for the test accuracy study. I mainly focused on information relating to the FC test result. In particular, I explored the units of measurements of FC levels in mmol/l and mg/kg in comparison to $\mu\text{g/g}$, which is described in Appendix B12. I also investigated FC test results with different operators (<, >, =) preceding numeric values of FC levels (Appendix B12).

I used the information gained from the descriptive statistics to formulate inclusion and exclusion criteria for FC tests into the test accuracy study. I compared included and excluded patients with FC tests using the two-sample t-test to compare continuous variables and the two proportions z-test to compare categorical variables.

4.3.3.2 Methods of analysis of test accuracy

4.3.3.2.1 Study design

The test accuracy study of FC testing for IBD used a retrospective cohort of adult patients ≥ 18 years with an FC test and no previous IBD diagnosis recorded in the THIN database. If patients had more than one FC test recorded only the first record was included.

Patients with FC tests included in the study could have had a preceding IBS or CRC diagnosis because these conditions do not exclude the possibility of a subsequent IBD diagnosis or FC testing. Patients could have had none, one, or more than one of the three diagnoses following their FC test. In cases where the patients had two diagnoses and the dates were different, I used the earlier diagnosis for the classification of the final disease status (IBD or non-IBD). While some may argue that the second diagnosis is more likely to be correct, the interval between the FC test and the second recording of the diagnosis might have been inappropriately long, leading to potential misclassification due to a change in disease severity. Where the dates were identical I chose IBD as the final diagnosis. As IBD and IBS can co-exist I based this decision on the assumption that, although these patients

were diagnosed with both conditions, IBD was the target condition of consideration for FC testing. I did not consider sensitivity analyses to test the assumption because the number of cases with two diagnoses in the test accuracy study was small (6/5,970 patients with FC tests).

4.3.3.2.2 Reference standard

The accepted reference standard for IBD is colonoscopy with histology.^{112, 158, 159} As secondary care interventions are under-coded in primary care, I used a composite reference standard to confirm the final diagnosis following testing. The reference standard was the diagnosis of IBD or a disease-specific medication in the form of a coded record in THIN. I considered three follow-up times of six months, 12 months and 24 months following testing to identify an IBD diagnosis. My aim was to explore the impact of a delayed diagnosis due to long waiting times for specialist care appointments or due to late recording of clinical codes. Only patients with FC tests with the complete follow-up time were included to prevent bias due to late diagnosis caused by insufficient follow-up time.

4.3.3.2.3 Outcomes

The main outcomes of the analysis were measures of test accuracy, including sensitivity, specificity, positive and negative predictive values. These were calculated in two ways: 1) People with a recorded IBD diagnosis were considered disease positive and people without IBD diagnosis (IBS, CRC and other) were disease negative. Disease negatives accounted for patients with an alternative diagnosis, those with no diagnosis and those with a previous IBS diagnosis prior to FC testing. 2) People with an IBD diagnosis were classed as disease positive and only people with an IBS diagnosis were classified as disease negative. The latter analysis allowed a comparison with hypothetical populations of IBD/IBS patients of published studies which formed the basis of the NICE guidance on FC testing.

4.3.3.2.4 ROC curves

I developed ROC curves to display the accuracy of FC testing at different thresholds for a positive test. I plotted ROC curves using two different datasets which take into consideration alternative ways the test results were recorded; 1) FC level preceded by any operator and 2) FC level preceded by an “=” operator. The first ROC curve considered all FC results but was restricted to thresholds between 33 and 300µg/g (see Appendix B12 for the rationale). Values smaller than or equal to 33µg/g were

collectively considered as at a threshold of 33µg/g and results greater than 300µg/g were all considered at the threshold 300µg/g.

The second ROC plot only included numeric results with the “=” operator. This plot has the advantage of summarising the complete range of positivity thresholds but is restricted to a subset of data.

I plotted ROC curves considering FC tests with six, 12 and 24 months of data available after the test date, respectively, to explore the impact of different follow-up times for IBD diagnoses on the complete range of sensitivity and specificity pairs. I used this exploratory analysis to decide on the follow-up time for the primary analysis and for subsequent analyses to compare thresholds (section 4.3.3.2.5) and contrast clinical questions (section 4.3.3.2.6), as well as for sensitivity analyses (section 4.3.3.2.9).

I compared ROC curves for different follow-up times using the method by Venkatraman for two unpaired ROC curves.¹⁶⁰ The assumption that the curves are not correlated might not hold. However, a paired comparison was not possible because data were of different sizes. The comparison provided an indication of any substantial differences between curves.

4.3.3.2.5 Accuracy at clinically relevant thresholds

I calculated measures of test accuracy at two different FC thresholds, namely 50 and 100µg/g. 50µg/g is the most commonly used threshold to interpret FC test results in published studies (section 4.2) and in clinical practice (Table 4.3). This threshold was recommended by manufacturers and adopted by test accuracy studies in secondary care. Following initial research into the use of FC testing in primary care the 100µg/g threshold was proposed by clinicians and subsequently endorsed by NICE for use in primary care with the aim to reduce the number of referrals of FC negatives to specialist care.¹¹⁸ I explored the effect of this change in threshold on the number of patients with IBD diagnosis identified by plotting the proportion of IBD diagnoses against the FC measurement.

4.3.3.2.6 Comparison of clinical questions

The FC test classifies tested people at the FC threshold into those with the target condition (IBD) and those without the target condition, which allows us to estimate

the test's ability to classify tested patients into those with and without IBD. A negative test suggests that IBD is absent rather than that the tested person has a specific alternative condition. Many previous test accuracy studies restricted the analysis to a pool of patients with either IBD or IBS diagnoses, where a negative test enabled the conclusion of IBS (see Appendix B9 for a list of studies). However, in primary care the non-IBD group is broader, comprising a range of alternative conditions. I investigated both scenarios and their effect on test accuracy measures to assess the applicability of findings when restricting the population to IBD and IBS patients.

4.3.3.2.7 Exploring non-IBD cases

I can be reasonably certain that a patient with a coded IBD diagnosis or IBD medication has IBD. However, the opposite does not apply to individuals without an IBD record. The non-IBD cases are made up of patients with a new IBS record, a former IBS record, a CRC record, any alternative diagnosis not considered a study variable, a missing Read code, or no diagnosis. A missed IBD record or an IBD diagnosis incorrectly coded results in incorrect classifications. If the number is substantial this could affect both sensitivity and specificity. To assess the size of the potential problem of missing IBD cases, I summarised the proportions of tested patients with and without IBD/IBS/CRC diagnosis by FC test outcome at six months after FC testing and >6 months allowing for late diagnoses. I further explored whether my data included any variables that were strong predictors of an IBD diagnosis, which could be used to identify potential missed cases. Variables included the two main co-morbidities of IBD and IBS, referral, and symptoms. I used univariable logistic regression analyses to quantify the strength of association between IBD diagnosis and the listed variables by producing odds ratios with 95% confidence intervals.

4.3.3.2.8 Time to diagnosis

With no specific test to diagnose IBS, recording of an IBS diagnosis may be substantially later than an organic condition such as IBD and CRC. Delayed diagnosis of IBS may suggest that non-IBD cases without diagnosis within six months of the FC test may be late IBS cases. This may underestimate the specificity in the differential diagnosis of IBD versus IBS when FC negative cases with a delayed IBS diagnosis were not included in the analysis. I plotted the time to diagnosis for IBD, IBS, and CRC to investigate whether IBS diagnoses were

recorded considerably later than organic conditions. I ignored diagnoses of IBS and CRC prior to the FC test. The time variable was the time from FC test to diagnosis or to the patient's end date if no diagnosis was recorded. I used Kaplan-Meier survival analysis methods and compared survival curves using the Log-rank test.

4.3.3.2.9 Sensitivity analyses

I undertook the following sensitivity analyses to investigate the effects that different assumptions have on the test accuracy estimates. I also explored the effects of excluding patients not eligible for testing as according to NICE guidance.

a) Sensitivity analyses regarding units of measurement

The primary analysis only included patients with test results reported in $\mu\text{g/g}$ because I could not ascertain that other units were used erroneously (Appendix B12). However, this excluded a proportion of patients which might be systematically different. This sensitivity analysis explored the impact of the decision to base the primary analysis on a subset of patients and to what extent the results differ when all patients with a numeric test result are considered. I did this in three analyses:

1. Units are equivalent: all numeric results included irrespective of the units of measurement (assumption that numeric results were measured in $\mu\text{g/g}$ but reported in different units and no conversion attempts were made)
2. Results in units other than $\mu\text{g/g}$ are all FC positive
3. Results in units other than $\mu\text{g/g}$ are all FC negative

b) Considering qualitative results with recorded upper reference range

The primary analysis only considered FC tested patients with numeric test results. While a qualitative result such as 'abnormal' might be sufficient for clinical decisions, qualitative results are uninformative for test accuracy at given thresholds. However, this meant that 1,887 patients with a qualitative FC test result were disregarded. This sensitivity analysis investigated the effect of including patients with a qualitative test result. These could be classified at the $50\mu\text{g/g}$ FC threshold in the presence of an upper reference range for interpretation (see Appendix B13 for a definition of the upper reference range).

c) Reference standard is Read code for IBD only

The primary analysis defined patients with an IBD Read code or an IBD treatment as having IBD. Patients with an IBD treatment recorded but without IBD Read code

may have been classified incorrectly as IBD patients. In this sensitivity analysis, I excluded cases of IBD classified solely on the basis of an IBD drug code and compared resulting test accuracy measures with those of the main analysis.

d) FC tests with subsequent referral or colonoscopy only

The primary analysis of test accuracy assumed that patients with an IBD diagnosis were seen in secondary care and received a diagnosis following some form of confirmatory investigation. This assumption is based on the knowledge that IBD is not generally diagnosed in primary care as it requires colonoscopy to assess extent and location of the inflammation. Without a record, however, this cannot be guaranteed. This sensitivity analysis focused on FC tested patients who were more likely to have received the preferred reference standard because 1) they were referred to secondary care within six weeks of the FC test or 2) they had a record of colonoscopy or sigmoidoscopy within one year of the FC test.

e) FC tests with prior NICE eligible symptoms only

The primary analysis included all eligible patients with an FC test regardless of symptoms recorded prior to testing. This could have resulted in selection bias. A test accuracy study should ideally recruit all patients with symptoms suggestive of IBD. NICE guidance restricts the indication for FC testing to patients with abdominal symptoms for more than six weeks when colorectal cancer is not suspected. This sensitivity analysis explored exclusion of patients not meeting the requirements in the NICE guidance: >50 years of age, symptoms lasting less than six weeks, or no symptoms recorded.

4.3.3.2.10 Impact of test results on referral rates

I computed proportions of tested patients referred with a true positive, false positive, false negative, and true negative test result in order to investigate the effect that the FC test results have on patient management. If the test results had no impact on referral, I would expect the proportions across groups to be similar. If GPs act on the test result (refer positives and manage negatives), the proportions of patients referred with true positive and false positive results should be 1) close to one; 2) equal; and 3) greater than those for false negative and true negative results (both close to 0). I compared the proportions of referred patients using the two proportions z-test.

4.3.3.2.11 Comparison of study outcomes with published results of test accuracy

I collated and summarised test accuracy measures from published primary care test accuracy studies and meta-analysed estimates of sensitivity and specificity from section 4.2 and compared them with my test accuracy estimates. I selected the meta-analysis for comparison that most closely resembled the THIN data considering threshold and clinical question. The meta-analysis included different testing assays while the test assays used in laboratories informing the THIN data are unknown. However, variation across laboratories and within laboratories over time is expected, therefore, resembling the scenario of the meta-analysis of different test assays.

4.3.4 RESULTS

4.3.4.1 Results of data management and inclusion criteria for FC tested patients

The THIN database contained information on 17,466 first time FC tested patients in a total of 6,965,853 adults between January 2006 and December 2016. Of these, 2,126 tests were recorded outside the patients' observation period and were excluded.

4.3.4.1.1 Exploration of FC test results recorded in THIN

In the THIN dataset, FC tests were reported as dates at which the FC test was undertaken. Each FC test came with a wealth of information that related to the result of the FC test. Table 4.3 provides an overview of the THIN variables that relate to the result of an FC test and Appendix B13 contains a description of the variables. Inspection of the information revealed that 1) test results were reported inconsistently; 2) numeric results were only reported for 9,500/17,466 (54.4%) FC tests; and 3) not all numeric results were reported with a unit of measurement and operator. Furthermore, there were a number of units and operators reported without corresponding numeric results.

Table 4.3 shows that $\mu\text{g/g}$ was the most commonly used unit to report FC levels. Following my exploration of FC test results with different units of measurement (Appendix B12), I was not confident that I could use units interchangeably and had to assume that units were truly different.

Table 4.3 Information available from THIN variables for 17,466 FC tests recorded in THIN relating to the FC test result

Medcode description*		Operator		Numeric result for FC		Units of measurement		Significance		Upper reference range	
Medcode	Records N	Operator	Records N	Numeric result	Records N	Unit	Records N	Significance level	Records N	Threshold	Records N
Calprotectin level	11339			Recorded	9500	#/[tot]	1	Abnormal	1820	0	9
FC content	5978	<	3100	NA	7966	%	14	Above high reference limit	137	20	5
FC test indeterminate	3	<=	3	Total	17466	g	10	High	3006	29	116
FC test invalid	1	=	7864			mg/g(dry wt)	13	Low	2	30	1
FC test negative	32	>	356			mg/kg	629	Normal	288	40	485
FC test positive	113	>=	1			mmol/L	841	Outside ref range	3	49	232
Total	17466	NA	6142			ng/mL	2	Potential abnormal	20	50	8985
		Total	17466			null value	13	Potentially abnormal	11	51	888
						u	60	Very High	8	55	2
						u/g	91	Very Low	1	59	19
						µg	2	NA	12170	60	1898
						µg/g	8137	Total	17466	70	38
						µg/g(dry wt)	33			111	6
						µg/g(wet wt)	46			112	74
						µg/L	7			2500	1
						NA	7567			2842	4
						Total	17466			NA	4703
										Total	17466

*See Appendix B13 for an explanation of the column headings
 FC faecal calprotectin, N total number, NA not available, wt weight

My exploration of operators of numeric FC test results (Appendix 12) revealed that a number of laboratories reported FC test results at a common threshold. Except for one value of $>30\mu\text{g/g}$, results reported with the operator “>” used the threshold of $300\mu\text{g/g}$ or above. This means all actual results, with this one exception, were greater than 50 and $100\mu\text{g/g}$. All results reported using the operator “<” used the threshold $33\mu\text{g/g}$ or smaller. This means all actual results were smaller than 50 and $100\mu\text{g/g}$.

4.3.4.1.2 Consideration of the insights into the reporting of test results in the inclusion criteria for FC tests into the test accuracy study

In terms of test accuracy, only numeric results with a valid unit are meaningful. For the primary analysis I, therefore, only included FC tested patients with a numeric test result in $\mu\text{g/g}$ to ensure findings were meaningful and valid. Patients who had an FC test without results, with invalid results, qualitative results only, with no units, or units other than $\mu\text{g/g}$ were excluded from the analysis. The flow diagram in Figure 4.7 describes the number of patients excluded for each reason. There is a risk of bias if the number of excluded patients with inappropriate test results is large and they do not resemble the included population. Table 4.4 compares the included and excluded population and in section 4.3.4.2.7 I address the decision in a number of sensitivity analyses to assess the level of risk.

At the thresholds of interest ($50\mu\text{g/g}$ and $100\mu\text{g/g}$) the operators > and < have no implications for the estimation of sensitivity and specificity. Only the one numeric result reported as $>30\mu\text{g/g}$ (Appendix B12) had an ambiguous interpretation at the two thresholds and was, therefore, excluded from the analysis. However, ROC curve analyses which calculate and graph sensitivity and specificity pairs at all available FC thresholds were likely to be affected by the considerable numbers of results reported as <33 and $>300\mu\text{g/g}$. To account for that, I plotted two alternative ROC curves, which I described in section 4.3.3.2.4.

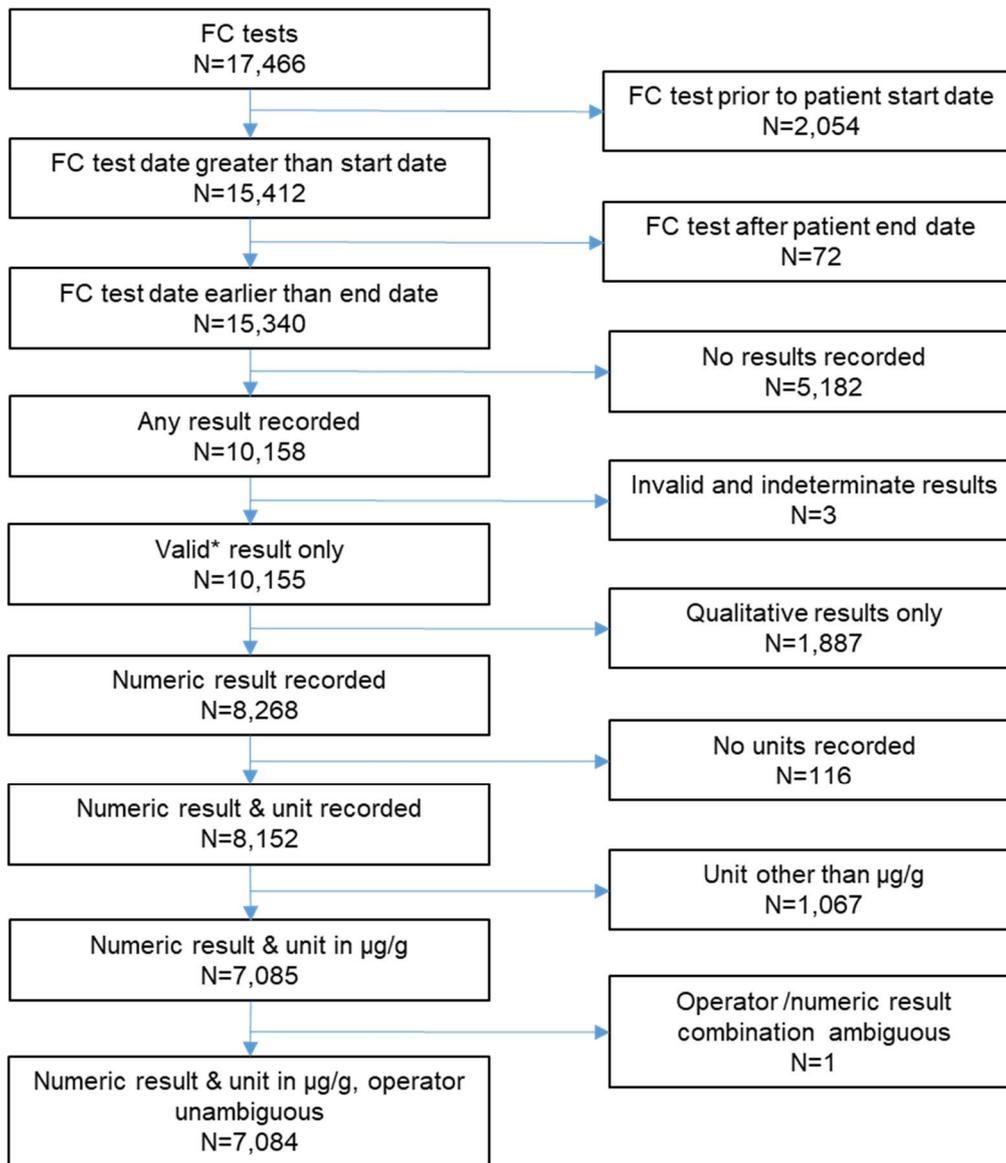


Figure 4.7 Flow diagram of inclusion criteria of FC tests into test accuracy study
*Results not coded invalid or indeterminate in THIN

4.3.4.2 Results of test accuracy

4.3.4.2.1 Characterisation of FC tested patients included into the test accuracy study

7,084/15,340 FC tested patients with an FC test recorded in the observation period were considered for the analyses (Figure 4.7). A comparison of the included and excluded patient groups showed regional differences, with more FC tests excluded from Wales, Northern Ireland and England and fewer excluded from Scotland (Table 4.4). There were also more excluded from the least deprived group according to Townsend Score, and fewer excluded from the most deprived group. The proportion of patients with an IBD or IBS diagnosis were slightly higher in the included than the excluded group.

Table 4.4 Patient characteristics of FC tested patients with appropriate test result (included) and inappropriate test result (excluded)

	Included (N=7,084)	Excluded (N=8,256)	p-value
Age (mean (SD)) at FC test date, range	44.07 (16.34), 18-101	43.33672 (16.49), 18-97	0.006
Sex (female (%))	4544 (64.1)	5247 (63.6)	0.521
BMI (mean (SD))	26.76674 (6.20)	27.01912 (6.26)	0.018
Not reported n (%)	788 (11.1)	929 (11.2)	
Ethnicity n (%)			
White	4,095 (57.8)	4,853 (58.8)	0.222
Asian	122 (1.7)	220 (2.7)	<0.0001
Black	81 (1.1)	117 (1.4)	0.134
Mixed	34 (0.5)	54 (0.7)	0.155
Other	45 (0.6)	43 (0.5)	0.350
Not reported	2,707 (38.2)	2,969 (36.0)	0.004
Townsend score n (%)			
1 (least deprived)	1,156 (16.3)	1,768 (21.4)	<0.0001
2	1,300 (18.4)	1,492 (18.1)	0.655
3	1,382 (19.5)	1,487 (18.0)	0.018
4	1,343 (19.0)	1,458 (17.7)	0.038
5 (most deprived)	1,234 (17.4)	977 (11.8)	<0.0001
Missing	669 (9.4)	1,083 (13.1)	<0.0001
Country (n) (%)			
England	4,596 (64.9)	6,226 (75.4)	<0.0001
Northern Ireland	16 (0.2)	565 (6.8)	<0.0001
Scotland	2174 (30.7)	749 (9.1)	<0.0001
Wales	298 (4.2)	716 (8.7)	<0.0001
CRC ever recorded n (%)	31 (0.4)	45 (0.5)	0.3447
IBS ever recorded n (%)	2,167 (30.6)	2,379 (28.8)	0.0164
IBD ever recorded n (%)	230 (3.2)	186 (2.3)	0.0002
IBD and IBS recorded n (%)	84 (1.2)	57 (0.7)	0.0014
IBD and CRC recorded n (%)	2 (0.03)	0 (0)	0.1268
No diagnosis recorded n (%)	4,484 (63.3)	5,532 (67.0)	<0.0001

BMI body mass index, CRC colorectal cancer, IBD inflammatory bowel disease, IBS irritable bowel syndrome, N total number, n number, SD standard deviation

The overview in Figure 4.8 illustrates the timing of recording of a diagnosis relative to the date of FC testing in the included population. 1,669/7,084 (23.6%) patients had an IBS diagnosis recorded at the time of FC testing and eight had a CRC diagnosis at the time of testing. Of those without prior diagnosis of IBD, IBS, or CRC (n=5,407), 823 had a single diagnosis of either IBD, IBS or CRC recorded following FC testing, 13 had an IBD as well as an IBS diagnosis recorded, and one patient had an IBD and a CRC diagnosis recorded. The majority of patients (n=4,570) had no record of any of the three conditions following their FC test.

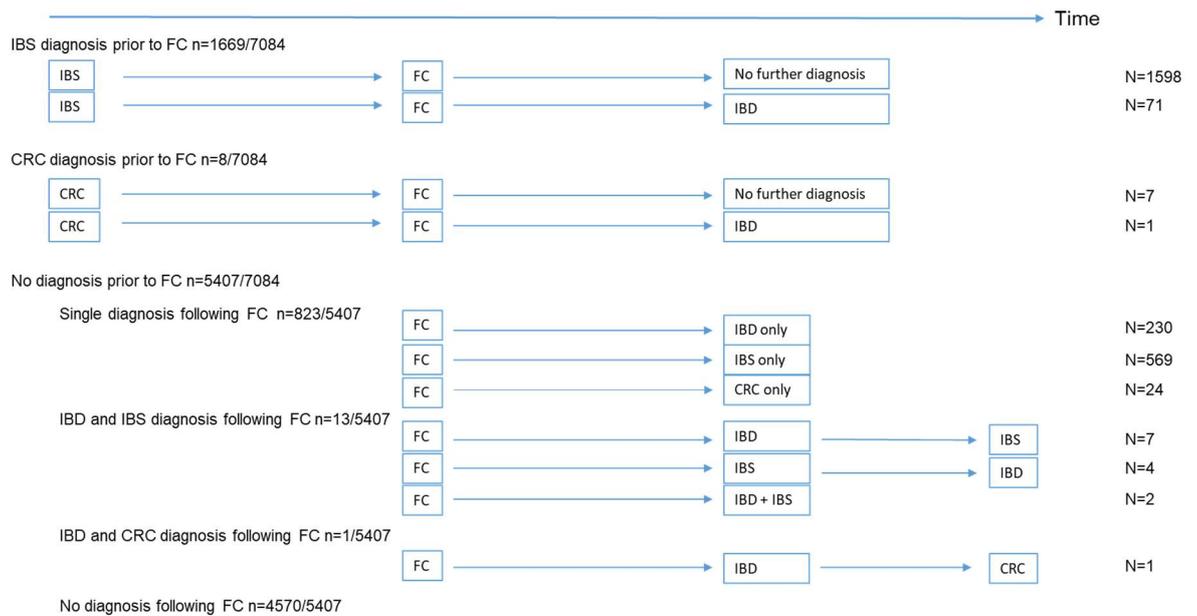


Figure 4.8 Overview of diagnoses prior and following FC testing for 7,084 included FC tested patients

4.3.4.2.2 Test accuracy of FC testing for IBD

This analysis considered 7,084 patients with an FC test. The target condition was IBD. Disease negatives were defined as not having IBD (including IBS, CRC or no record of any of the three conditions). Only FC tested patients with follow-up data of six months, 12 months and 24 months available following FC testing were included in the respective analyses. The resulting datasets had the following sizes: 5,970 patients with six months follow-up, 4,793 patients with 12 months follow-up and 2,662 patients with 24 months follow-up (Figure 4.9).

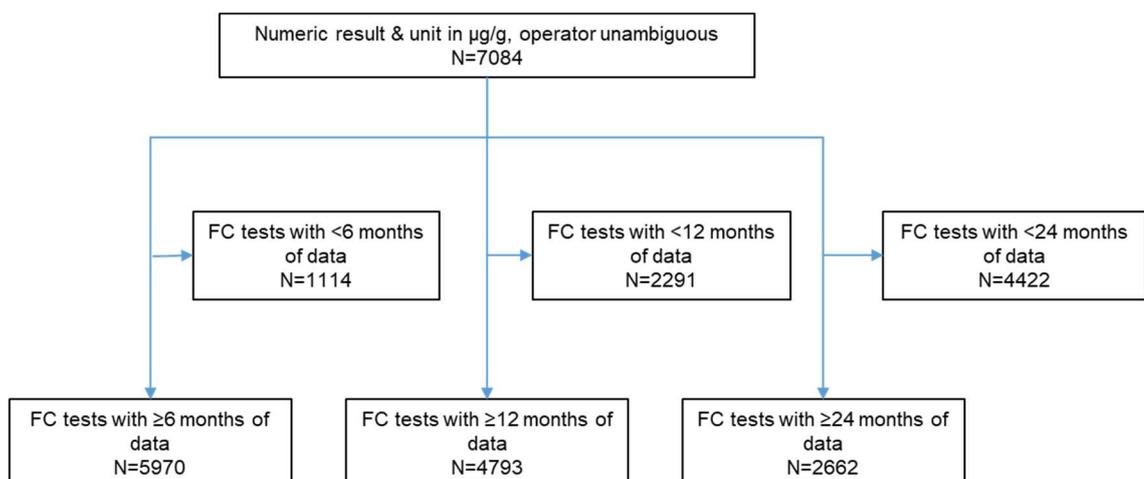


Figure 4.9 Datasets of FC tested patients with three different follow-up times available for three analyses

The sensitivity of FC testing to detect IBD at a threshold of 50µg/g was 0.93 (95% CI 0.89 to 0.96) in the 5,970 patients who could be followed up for at least six months after the FC test date. Specificity was 0.61 (95% CI 0.6 to 0.63). The positive predictive value was very low (0.08, 95% CI 0.07 to 0.09) because IBD is a rare condition and the test is not highly specific. The negative predictive value was close to perfect (0.996, 95% CI 0.993 to 0.997) because of the low IBD prevalence. Restricting the dataset to patients with a minimum of 12 months or 24 months follow-up reduced the dataset by about 20% and 55%, respectively (Figure 4.9). This resulted in a selected population with concerns over its generalisability. Considering additional IBD cases up to 12 and 24 months following testing had no impact on the measures of test accuracy (Table 4.5). Small changes observed in the point estimates may be due to the different population size and the different make-up of the study population.

Table 4.5 Test accuracy measures of FC testing at threshold of 50µg/g

Time*	N	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
6 months	5,970	195	2,219	15	3,541	0.93 (0.89 to 0.96)	0.61 (0.6 to 0.63)	0.08 (0.07 to 0.09)	0.996 (0.993 to 0.997)
12 months	4,793	198	1,790	16	2,789	0.93 (0.88 to 0.95)	0.61 (0.59 to 0.62)	0.1 (0.09 to 0.11)	0.994 (0.991 to 0.997)
24 months	2,662	129	1,067	11	1,455	0.92 (0.86 to 0.96)	0.58 (0.56 to 0.6)	0.11 (0.09 to 0.13)	0.993 (0.987 to 0.996)

*Time following FC testing to look for a record of IBD

CI confidence interval, FN false negatives, FP false positives, N sample size, NPV negative predictive value, PPV positive predictive value, TN true negatives, TP true positives

The ROC curves in Figure 4.10 describe the sensitivity and false positive rate (1-specificity) for the three different follow-up times at additional thresholds. The 50µg/g threshold prioritises sensitivity over specificity in all three analyses and keeps false negatives at a minimum. There was generally good agreement between the three different follow-up times for the range of thresholds from 33-300µg/g (E=0.0017, p=0.406, comparing six months follow-up with 24 months follow-up).

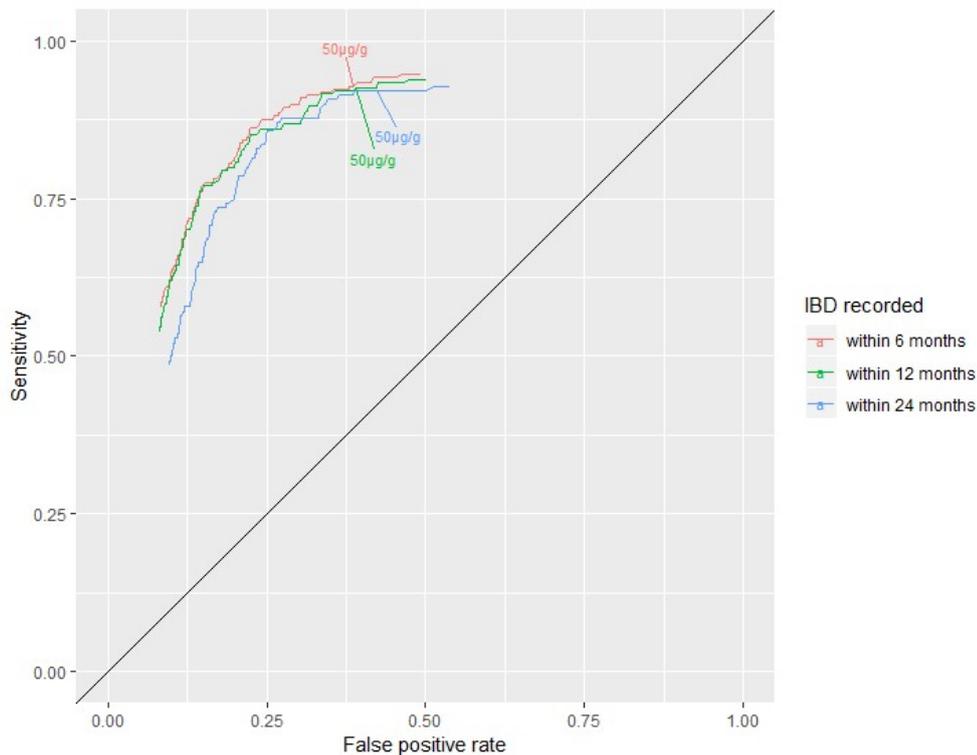


Figure 4.10 ROC curve of sensitivity and false positive rate (1-specificity) for thresholds 33-300µg/g for the differential IBD versus non-IBD for tests with at least six months (N=5,970), 12 months (N=4,793) and 24 months (N=2,662) follow-up available after the FC test date.

Restricting my analyses to patients who had an FC test result reported with the operator “=” only, reduced the number of patients to 4,549 (Figure 4.11). ROC curves drawn from this dataset included sensitivity and specificity pairs for all available thresholds (Figure 4.12). The position of the ROC curve may suggest a change in sensitivity and specificity for some thresholds for this reduced dataset. Comparison of ROC curves over the complete range of sensitivity and specificity suggests again that longer follow-up did not result in considerable numbers of additional IBD cases ($E=0.0017$, $p=0.613$, comparing six months follow-up with 24 months follow-up).

I, therefore, restricted subsequent analyses to the largest dataset of patients with six months follow-up.

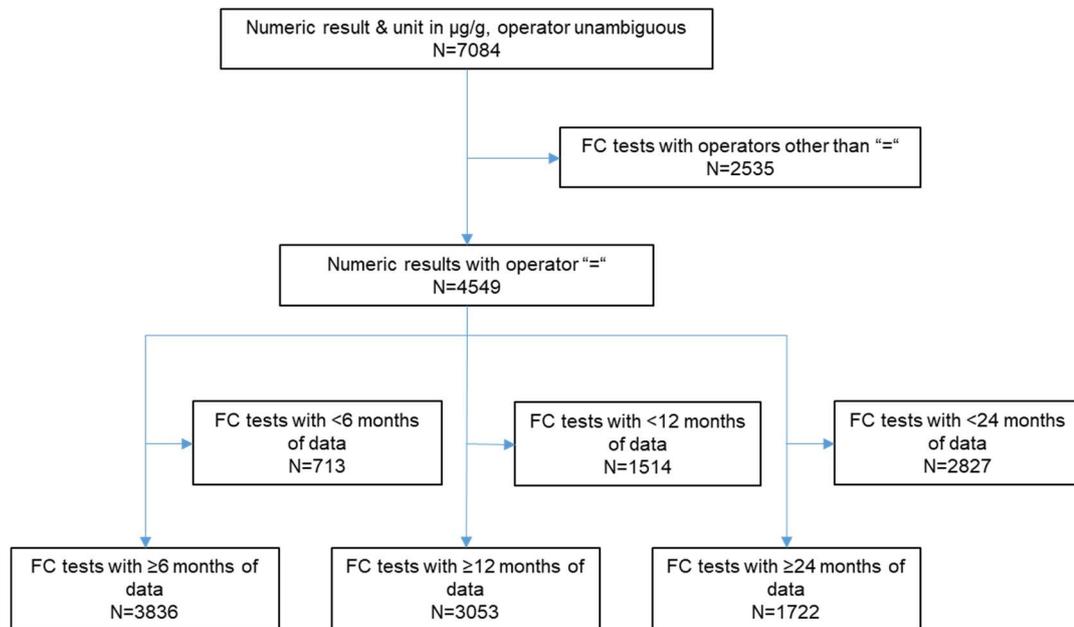


Figure 4.11 Datasets for three follow-up times of patients with numeric FC test results preceded by “=”

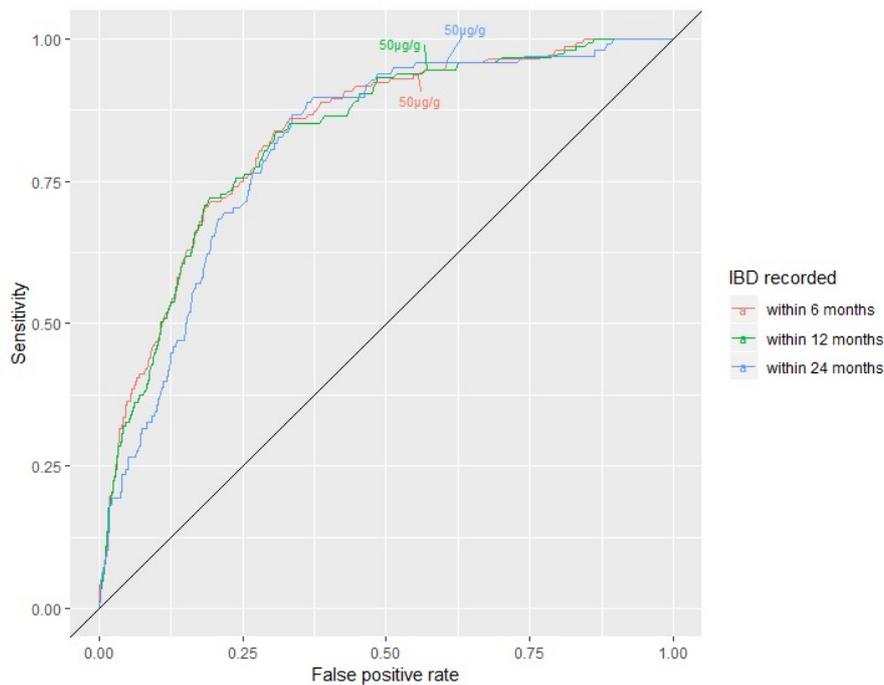


Figure 4.12 ROC curve of sensitivity and false positive rate (1-specificity) for all thresholds considering test results reported with operator “=” for the differential IBD versus non-IBD for patients with at least six months (N=3,836), 12 months (N=3,053) and 24 months (N=1,722) follow-up available after the FC test date.

4.3.4.2.3 Exploring thresholds

In Figure 4.13 I illustrate the interrelationship between sensitivity and specificity over a range of thresholds in an alternative plot. It clearly shows that the increase in threshold from 50 $\mu\text{g/g}$ to 100 $\mu\text{g/g}$ had a greater impact on specificity than on sensitivity. In numbers, this translated to an increase in specificity from 0.61 (95% CI 0.6 to 0.63) to 0.77 (95% CI 0.76 to 0.78) (Table 4.6). The impact on sensitivity was less compelling. The higher threshold reduced the number of false positives by 875 at the expense of missing an additional 14 cases of IBD in a population of 5,970 tested patients. The range of FC levels in question, in the debate of increasing the threshold from 50 to 100 $\mu\text{g/g}$, is narrow considering the whole spectrum of FC levels in patients with an IBD diagnosis (Figure 4.13). These data suggest that the proportion of IBD cases that may be missed by the higher threshold is less than 7% (14/210).

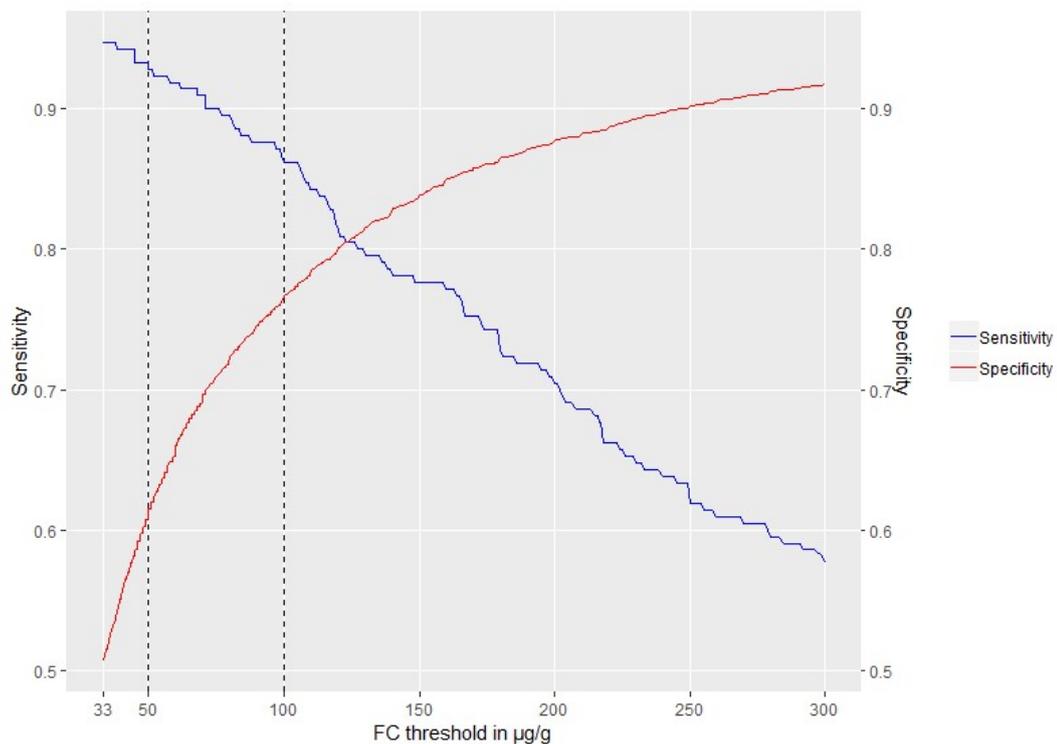


Figure 4.13 Sensitivity and specificity at FC thresholds 33-300 $\mu\text{g/g}$ for IBD versus non-IBD for patients with six months follow-up (N=5,970 any operator)

Table 4.6 Test accuracy measures of FC testing at threshold of 100µg/g compared to 50µg/g

Threshold	N	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
50µg/g	5970	195	2219	15	3541	0.93 (0.89 to 0.96)	0.61 (0.6 to 0.63)	0.08 (0.07 to 0.09)	0.996 (0.993 to 0.997)
100µg/g	5970	181	1344	29	4416	0.86 (0.81 to 0.9)	0.77 (0.76 to 0.78)	0.12 (0.1 to 0.14)	0.994 (0.991 to 0.996)

CI confidence interval, FN false negatives, FP false positives, N sample size, NPV negative predictive value, PPV positive predictive value, TN true negatives, TP true positives

4.3.4.2.4 Comparison of different clinical questions

Figure 4.14 illustrates the two datasets available for the comparison of FC test accuracy for two differentials: IBD versus non-IBD, and IBD versus IBS. The two datasets were of considerably different sizes. This was the result of excluding patients without an IBD or IBS record to create a patient population with a mix of IBD and IBS diagnoses only. The large number of exclusions stresses the limited applicability of the population to the clinical context.

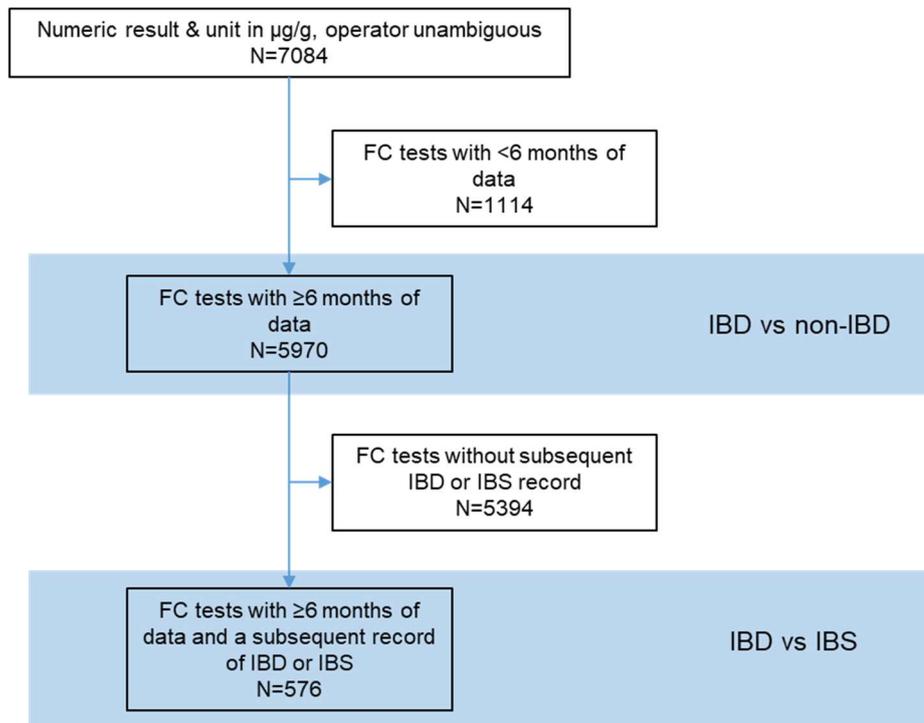


Figure 4.14 Illustration of the size of the available datasets for the comparison of test accuracy for the two differentials IBD versus non-IBD and IBD versus IBS

Changing the definition of the non-diseased group from non-IBD to IBS reduced the number of false positives and true negatives considerably. This led to an increase in point estimates of specificity by about 10% and in PPV by about 60% (Table 4.7) at the common threshold of 50µg/g. The ROC curves in Figure 4.15 depict the sensitivity and false positive rate (1-specificity) for both scenarios for all available thresholds.

Table 4.7 Test accuracy measures for FC testing at the threshold of 50µg/g for the detection of IBD recorded within six months of FC testing considering two different non-target conditions

	N	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
IBD vs non-IBD	5,970	195	2219	15	3541	0.93 (0.89 to 0.96)	0.61 (0.6 to 0.63)	0.08 (0.07 to 0.09)	0.996 (0.993 to 0.997)
IBD vs IBS	576	195	104	15	262	0.93 (0.89 to 0.96)	0.72 (0.67 to 0.76)	0.65 (0.6 to 0.7)	0.946 (0.913 to 0.967)

CI confidence interval, FN false negative, FP false positive, IBD inflammatory bowel disease, IBS irritable bowel syndrome, N sample size, NPV negative predictive value, PPV positive predictive value, TN true negative, TP true positive

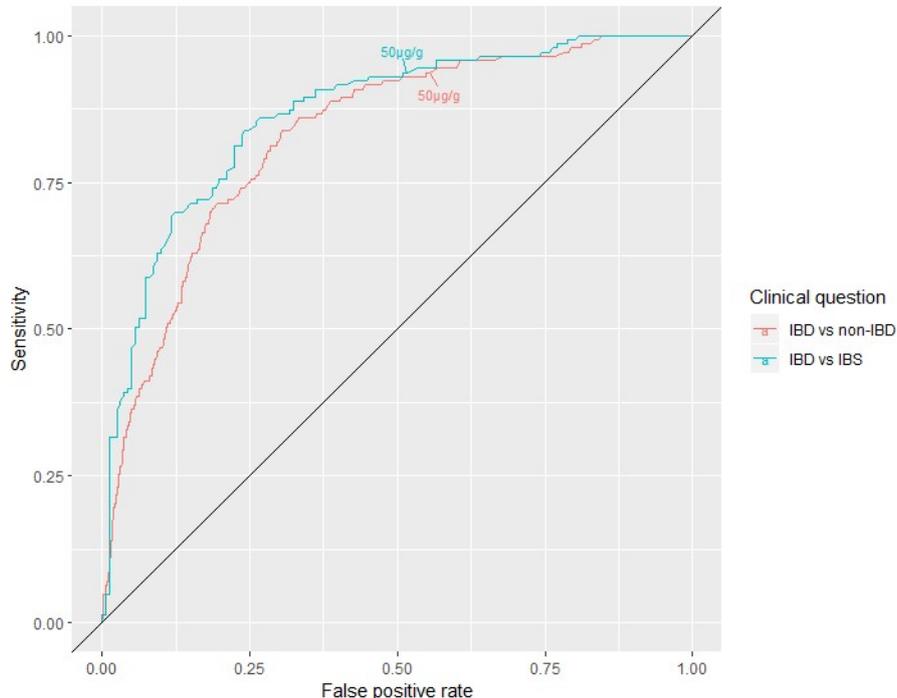


Figure 4.15 Comparison of ROC curves of FC testing for IBD versus non-IBD (n=3,836) and IBD versus IBS (n=304) considering FC test results recorded with the “=” operator

4.3.4.2.5 Exploration of non-IBD cases

The definition of disease negative patients included those with a new IBS record, a former IBS record, a CRC record, other conditions not considered as study variables, a missing Read code, or no diagnosis. An FC test without subsequent diagnosis is difficult to interpret without consideration of all alternative diagnoses. Figure 4.16 shows that the proportion of FC tests without subsequent diagnosis of IBD, IBS or CRC is large. 87% (2,096/2,414) of patients with a positive FC test result had no recorded diagnosis of IBD, IBS, or CRC six months after the FC test date. Of these, 501 had a previous diagnosis of IBS or CRC which would not be recorded again. At the end of the complete study period, there were still 1,470 (61%) test positive patients without a diagnosis of interest. Of test negative patients, 92% (3,277/3,556) had no diagnosis recorded at six months and 64% (2,284/3,556) had no diagnosis recorded at any time after the FC test date. My regression analysis to investigate the association of an IBD diagnosis with a number of study variables considered symptoms (e.g. diarrhoea and abdominal pain), co-morbidities (e.g. asthma and depression), referral and further FC testing. It identified no strong predictors of an IBD diagnosis on which basis to reliably reclassify non-IBD cases (Appendix B14). Therefore, for more than half of the included patients the final diagnosis was not available and their classification as IBD negative may be contentious. Furthermore, I could not quantify the potential bias caused by missing IBD diagnoses.

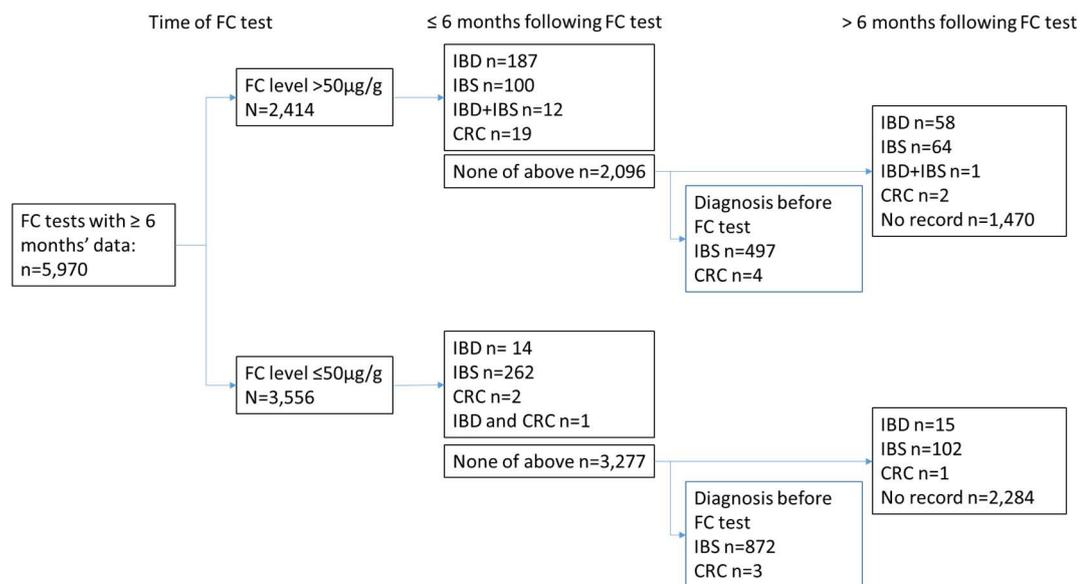


Figure 4.16 Proportions of patients with and without IBD/IBS/CRC diagnosis following FC testing by test outcome

4.3.4.2.6 Time to diagnosis

The Kaplan-Meier plot in Figure 4.17 compares the time to diagnosis for CRC, IBD, and IBS from the FC test date. I considered a diagnosis at any time following the FC test in the population of 7,084 FC tested patients. The graph shows that IBS diagnoses were recorded until nearly six years after the FC test, when a link to the initial FC test cannot be readily assumed. Time to diagnosis appeared slightly shorter for the majority of CRC cases. However, there was no statistically significant difference (Log rank test $p < 0.47$). More importantly, time to diagnosis of IBD and IBS was very similar within the first six months of the FC test. The majority of the individuals had their diagnosis during this time (75% for IBD and 70% for IBS). This indicates that choosing the six months follow-up to identify an IBD diagnosis did not appear to have biased the test accuracy results towards IBD diagnosis in the differential of IBD versus IBS.

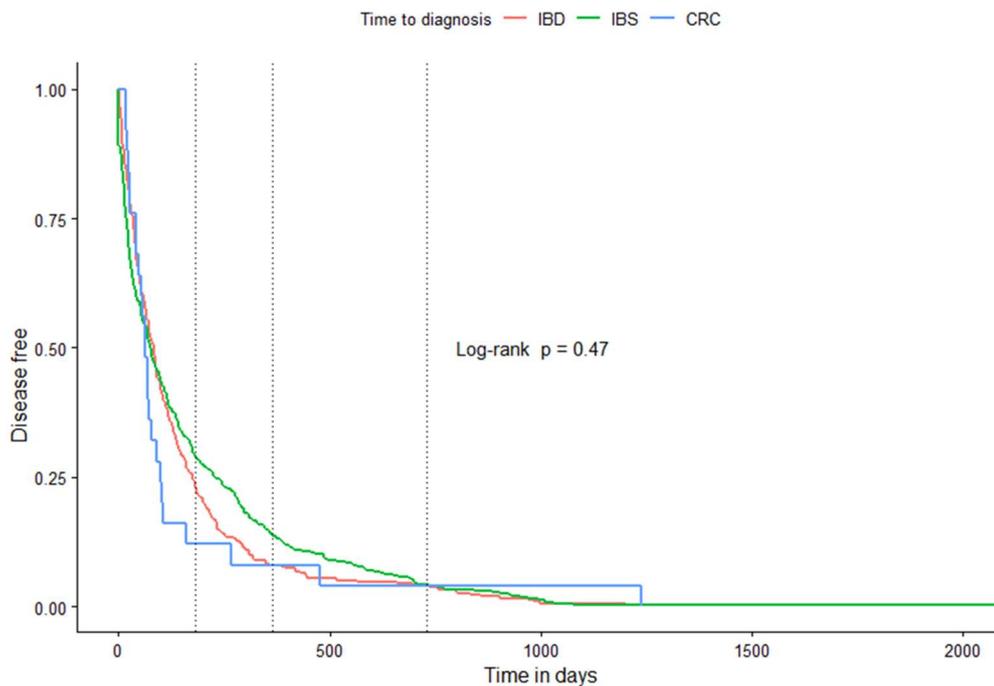


Figure 4.17 Kaplan Meier time to event curve depicting time to diagnosis for cases of IBD, IBS and CRC from the time of FC testing

4.3.4.2.7 Sensitivity analyses

My sensitivity analyses, which explored several assumptions of my test accuracy study, revealed overall greater uncertainty in the estimate of specificity than sensitivity (Table 4.8). Considering all numeric results irrespective of the units of measurement used added 788 FC tests to the analysis. However, they had no impact on the estimates of test accuracy. The extreme assumptions that the units of the 788 FC tests were truly different and would translate to either all positive or all

negative, if transformed into $\mu\text{g/g}$, would have decreased the specificity and increased the sensitivity in turn.

Interpretation of qualitative results as positive or negative at the $50\mu\text{g/g}$ threshold included an additional 1,030 FC tested patients in the analysis. However, the vast majority of results were positive contributing disproportionately to the false positives, therefore, reducing the specificity to 0.53 (95% CI 0.52 to 0.54).

Only 569 FC tests were ordered for patients who had symptoms recorded that fitted the NICE eligibility criteria. Test accuracy measures in this subset did not differ significantly from those estimated using the full dataset. The proportion of patients with NICE eligible symptoms was similar for true positives, false positives, false negatives and true negatives (close to 10% of the full dataset) (Table 4.8). This would suggest that abdominal symptoms, or the recording of such, was not associated with either a positive or negative FC test outcome.

The primary analysis also included 79 patients who had an IBD diagnosis affirmed by an IBD specific drug code only. Reclassifying these as IBD negative (Read code negative) had no significant impact on the test accuracy estimates (Table 4.8).

Of the 5,970 FC tested patients, 52% were referred and 22% had a colonoscopy scheduled. Table 4.8 reveals that predominantly true negatives were excluded from the datasets of referred and colonoscoped patients. Consequently, the specificities of FC testing in these subgroups were lower compared to the primary analysis (0.55, 95% CI 0.53 to 0.56 for the referred and 0.39, 95% CI 0.36 to 0.42 for the colonoscoped subgroup).

Table 4.8 Sensitivity analyses compared to primary analysis of FC testing for IBD vs non-IBD at a threshold of 50µg/g with IBD recorded within six months

Analysis	N	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
Primary analysis	5,970	195	2,219	15	3,541	0.93 (0.89 to 0.96)	0.61 (0.6 to 0.63)	0.08 (0.07 to 0.09)	0.996 (0.993 to 0.997)
FC tests with any unit	6,758	220	2,481	17	4,040	0.93 (0.89 to 0.95)	0.62 (0.61 to 0.63)	0.08 (0.07 to 0.09)	0.996 (0.993 to 0.997)
Numeric results other than µg/g all FC positive	6,758	222	2,980	15	3,541	0.94 (0.9 to 0.96)	0.54 (0.53 to 0.56)	0.07 (0.06 to 0.08)	0.996 (0.993 to 0.997)
Numeric results other than µg/g all FC negative	6,758	195	2,219	42	4,302	0.82 (0.77 to 0.87)	0.66 (0.65 to 0.67)	0.08 (0.07 to 0.09)	0.990 (0.987 to 0.997)
Qualitative results included*	7,000	261	3,177	15	3,547	0.95 (0.91 to 0.97)	0.53 (0.52 to 0.54)	0.08 (0.07 to 0.09)	0.996 (0.993 to 0.997)
NICE eligibility**	569	20	194	1	354	0.95 (0.77 to 0.99)	0.65 (0.61 to 0.68)	0.09 (0.06 to 0.14)	0.997 (0.984 to 0.9995)
Read code only	5,970	148	2,266	6	3,550	0.96 (0.92 to 0.98)	0.61 (0.6 to 0.62)	0.06 (0.05 to 0.07)	0.998 (0.996 to 0.9992)
Referred only (within 6 weeks)	3,078	138	1,330	11	1,599	0.93 (0.87 to 0.96)	0.55 (0.53 to 0.56)	0.09 (0.08 to 0.11)	0.993 (0.988 to 0.996)
Colonoscoped only	1,299	110	725	5	459	0.96 (0.90 to 0.98)	0.39 (0.36 to 0.42)	0.13 (0.11 to 0.16)	0.989 (0.975 to 0.995)

*With a record of an upper reference range of 50 or above

**Patients <50 years and abdominal symptoms for >6 weeks

FC faecal calprotectin, FN false negatives, FP false positives, NPV negative predictive value, PPV positive predictive value, TN true negatives, TP true positives

4.3.4.2.8 Impact of test results on referral decisions

Table 4.9 shows the classification of all included patients and those referred within six weeks of their FC test according to their test result and disease status. If GPs referred patients based on positive FC test results, they would have referred 2,414/5,970 (40.4%) included patients. This would have detected 195/210 (92.9%) IBD cases. However, six weeks after FC testing 3,078/5,970 (51.6%) patients were referred, including 1,610/3,556 (45.3%) FC negative patients and only 149/210 (71.0%) of IBD cases. Therefore, the test result was not the single determining factor for or against a timely referral. Patients with true positive and false negative results were equally likely to be referred for further tests (70.8% vs 73.3%, $z=2$, $p=0.0371$). Both groups were referred more frequently than patients with false positive (59.9%) ($z=3$, $p=0.0028$) or true negative results (45.3%) ($z=7$, $p<0.0001$).

Table 4.9 Patients referred by test result and disease status

	TP	FP	FN	TN	Total
Classification of FC tested patients in primary analysis n	195	2,219	15	3,541	5,970
Classification of FC tested patients with referral* n	138	1,330	11	1,599	3,078
Proportion of patients in primary analysis referred*, %	70.8	59.9	73.3	45.3	51.6

*referral record within six weeks of FC test

FC faecal calprotectin, FN false negative, FP false positive, TN true negative, TP true positive

This may suggest that GPs used alternative cues (in the history, examination or other tests) which raised their suspicion of IBD. Overall, GP assessment plus FC testing may be worse than FC testing alone. Compared to FC testing alone, the number of patients with a true positive result was lower for GP assessment plus testing because GPs did not refer all test positive patients within six weeks. GPs referred some patients with a false negative FC test result on the test, but not enough to make up for not referring patients with true positive results on the test.

This interpretation assumes that referrals were recorded consistently for those with and without IBD record and irrespective of FC test outcome. I interpret the findings with caution because I could not confirm the final classification of more than half of the tested patients (section 4.3.4.2.5) and missing IBD cases cannot be excluded. Furthermore, all IBD patients would have been eventually referred according to my definition of an IBD diagnosis which assumes that IBD diagnoses are only recorded

following confirmation by a gastroenterologist and invasive testing. 61/210 (29.0%) IBD cases were either referred after six weeks or their referral was not recorded.

4.3.4.2.9 Comparison of sensitivity and specificity to published estimates

Table 4.10 compares different estimates of sensitivity and specificity for FC testing to detect IBD versus non-IBD. Estimates include published estimates from studies in primary care, my meta-analytical result of primary and secondary care studies reported in section 4.2, and my estimates using the THIN data reported in section 4.3.4.2.2. At the 50µg/g threshold the sensitivity and specificity established with the THIN data sit within the range of estimates based on the published literature. However, primary care studies included small numbers of cases and presented estimates with large confidence intervals, which means they are not informative for a useful comparison. Considering both sensitivity and specificity, the THIN analysis aligns with results from the meta-analysis of 14 studies from primary and secondary care, judged by the overlapping confidence intervals. The definition of disease appears similar for IBD and non-IBD across the analyses as far as I can ascertain. Exclusion of microscopic colitis from the definition of IBD by Boyed et al. (2016)⁶¹ seems to be the main difference. However, the meta-analysis is based on studies with an overall greater prevalence of IBD than the studies based on purely primary care populations most likely because of the inclusion of secondary care studies.

Similar observations could be made for the 100µg/g threshold (Table 4.10). A comparison of the two thresholds showed that the sensitivity at the 100µg/g threshold was about seven percentage points lower compared to the sensitivity at the 50µg/g threshold based on the THIN data. It was 21 percentage points lower in the meta-analysis of five studies at the 100µg/g than in the meta-analysis of 14 studies at the 50µg/g threshold. This may suggest that proportionately more IBD cases with FC levels in the 50-100µg/g region existed in the published studies than in the THIN data. The difference in point estimates for specificity was 15 percentage points in the THIN data and 16 percentage points in the meta-analysis when comparing the two thresholds.

Similarly, THIN estimates for the differential of IBD versus IBS (sensitivity 0.93, 95% CI 0.89 to 0.96 and specificity 0.72, 95% CI 0.67 to 0.76) were comparable to estimates from the meta-analysis of 11 secondary care studies (sensitivity 0.96, 95% CI 0.92 to 0.98 and specificity 0.77, 95% CI 0.66 to 0.85) (section 4.2.4.1).

Table 4.10 Estimates of sensitivity and specificity for FC testing for IBD versus non-IBD based on the THIN analysis compared with published results from primary care and meta-analytical results of primary and secondary care studies at two thresholds (section 4.2.4)

	My THIN analysis	Studies from primary care			My MA of primary and secondary care studies*
		Conroy 2017 ⁶³	Hogberg 2017 ⁶²	Boyed 2016 ⁶¹	
Threshold 50µg/g					
Sensitivity (95% CI)	0.93 (0.89 to 0.96)	0.73 (0.43 to 0.90)	0.90 (0.60 to 0.98)	1.0 (0.86 to 1.0)	0.94 (0.90 to 0.97)
Specificity (95%CI)	0.61 (0.6 to 0.63)	0.65 (0.60 to 0.69)	0.73 (0.68 to 0.77)	0.71 (0.66 to 0.75)	0.67 (0.57 to 0.76)
Threshold 100µg/g					
Sensitivity (95% CI)	0.86 (0.81 to 0.9)	0.55 (0.28 to 0.79)	0.7 (0.40 to 0.89)	NR	0.73 (0.60 to 0.82)
Specificity (95%CI)	0.77 (0.76 to 0.78)	0.8 (0.76 to 0.84)	0.86 (0.82 to 0.89)	NR	0.82 (0.77 to 0.86)
Definition of disease					
Definition of IBD	Record of a Read or drug code suggestive of IBD including microscopic and indeterminate colitis	IBD including microscopic colitis	IBD including microscopic colitis and unspecific colitis	IBD excluding microscopic colitis	IBD***
Definition of non-IBD	No IBD record	No further investigation and non-IBD organic conditions**	No or any other pathology including CRC, HRA	Gastroenteritis, microscopic colitis and IBS	Non-IBD***
Disease prevalence					
IBD	210/5,970 (3.5%)	11/410 (2.7%)	10/384 (2.6%)	23/424 (5.4%)	342/2,993 (13%)

*Based on 14 studies for FC threshold 50µg/g and five studies for FC threshold 100µg/g (see Table 4.2)

**Non-IBD organic conditions included CRC, colorectal adenomatous polyps, diverticulitis, appendicitis and diversion proctitis

***As defined by heterogeneous studies

CI confidence interval, CRC colorectal cancer, HRA high-risk adenoma, IBD inflammatory bowel disease, IBS irritable bowel syndrome, MA meta-analysis, NR not reported, THIN the Health Improvement Network

4.3.5 DISCUSSION

4.3.5.1 Study results

The sensitivity of FC testing to detect IBD in routine primary care was high (0.93, 95% CI 0.89 to 0.96), but the specificity was modest (0.61, 95% CI 0.6 to 0.63). Moreover, the sensitivity analyses demonstrated greater uncertainty in my estimates of the specificity. These findings were based on a selected population of FC tested patients who had an FC level recorded in the THIN database. The high number of false positives may be irrelevant in clinical practice if the group represents alternative inflammatory conditions that warrant further assessment. Restricting the group without IBD to those with a confirmed IBS record increased the specificity by 11 percentage points but has little meaning for clinical practice because it is an artificial patient population. Extending the follow-up time to allow for late recordings of IBD had no impact on sensitivity and specificity. The ROC analysis revealed that the recommended FC threshold of 50µg/g prioritises sensitivity over specificity.

Changing the threshold from 50µg/g to 100µg/g had a greater effect on the specificity than the sensitivity. For clinical practice this could translate into fewer referrals of non-IBD patients to gastroenterology at the expense of missing an additional nearly 7% of IBD cases. This may be acceptable if two conditions apply. Firstly, the non-IBD cases who are not referred have non-inflammatory disease that can be managed in primary care. Secondly, missed IBD patients have milder disease and would return with a flare of symptoms within a few months of clinical suspicion.¹⁶¹ This interpretation is purely based on the expectation that GPs refer FC positive patients and manage FC negative patients. However, investigation of the number of referrals of true positive, false positive, false negative, and true negative patients showed that the actual management of tested patients did not follow this expected behaviour. The impact of testing on the management decision, especially in cases of negative tests where clinical uncertainty remains, is of interest and is explored in Chapter 5.

It is important to consider the generalisability of the test accuracy estimates in light of the study assumptions and the great proportion of excluded patients from the study. These limitations need to be weighed against the strengths of the study design. In section 4.3.5.2 I discuss the strengths of the THIN test accuracy study as well as the limitations under the headings: selection bias (section 4.3.5.2.1), missing data (section 4.3.5.2.2), reference standard (section 4.3.5.2.3), and confirming disease status in non-IBD cases (section 4.3.5.2.4). In section 4.3.5.3 I discuss the

findings in light of published evidence. In section 4.3.5.4 I discuss the suitability of the THIN database for test accuracy studies in general, and list criteria under which routine data for test accuracy is useful. I further appraise my THIN test accuracy study against these criteria in order to judge the validity of my study findings.

4.3.5.2 Strengths and limitations

A particular strength of using routine data to estimate a test's accuracy is that the sample size is likely to be larger than the usual cross-sectional studies used to evaluate a test. This is particularly useful for target disorders such as IBD which have a low prevalence. It enabled me to produce point estimates with narrow statistical confidence intervals. This is in contrast to other primary care studies.⁶¹⁻⁶³

I applied strict eligibility criteria to ensure the reliability of reported test accuracy measures. However, this meant a trade off in generalisability because the excluded population may be different to the included population beyond what I was able to test.

I explored the effects of different follow-up times on the definition of IBD disease, which allowed me to choose the most appropriate follow-up. Finally, I undertook a number of sensitivity analyses, which tested my main assumptions during data management. I could show that the sensitivity for FC testing varied little across the analyses.

Using routine data for test accuracy is based on a number of assumptions and limitations, which raises the question as to what extent we can believe the results. The main limitations important to my research are 1) selection bias, 2) missing data mainly affecting the FC test results, 3) the reference standard, and 4) the inability to confirm absence of disease. Misclassification through miscoding using IBD sub-codes is a negligible problem, which I explained in section 3.6.2.4. Multiple disease codes were of little concern to the study as the number of patients with both an IBD and an IBS code was negligible, with only six cases in the test accuracy study.

4.3.5.2.1 Selection bias

A limitation of the test accuracy study was the inclusion of patients with a recorded FC test rather than symptoms. Oostenbrink et al. (2003) argued that studies suffer from selection bias if not all patients with symptoms suspicious of disease are

considered in a test accuracy study.¹⁴⁹ However, the Venn diagram in Figure 3.10 showed that there was very little overlap between patients with NICE eligible symptoms and those with an FC test record. My sensitivity analysis showed that focusing on patients with eligible symptoms who had a subsequent FC test recorded would have drastically reduced the dataset by 90%. However, considering point estimates and confidence intervals, test accuracy was similar in this patient population. Selection bias may have been minimal.

4.3.5.2.2 Missing data

Close to half of the FC tested patients had no numeric test result. These tests are uninformative for test accuracy and I excluded FC tested patients without test results. This might have biased my results¹⁶² because missing data in routine care tend not to occur completely at random.¹⁴⁹ If positive test results were more reliably recorded than negative results, exclusion of these tests could have underestimated specificity and/or overestimated sensitivity. However, as FC testing is a recent test and electronic reporting of results can be expected I expect this to be minimal.¹¹⁰ Excluded tests were statistically different in some variables tested. I would expect that because the large dataset meant that the tests were highly powered. The clinical significance of the established differences is uncertain.

Although imputation methods are often used to mitigate against missing data I did not implement them for three reasons. Firstly, in observational data, observed variables are not sufficient to account for the differences between missing and observed values.¹⁶² Secondly, for imputation methods to be meaningful the proportion of missing data should be small and the missing subpopulation should be sufficiently similar to the wider population, neither of which could be confirmed. Thirdly, in order to build reliable imputation models, more meaningful variables are needed including severity of symptoms, assay type, referral to gastroenterology and information on disease verification method, none of which is available in the THIN database. Therefore, I could only address potential biases by sensitivity analyses.¹⁶²

4.3.5.2.3 Reference standard

Determination of test accuracy relies on measuring the level of agreement between the FC test results and the reference standard.¹⁶³ The reference standard determines the true disease state of tested patients in order to verify the FC test outcomes. The assessment of test accuracy, therefore, relies on: 1) the reference

standard being infallible; 2) the index test and reference standard being undertaken in a time interval which ensures that they measure the same disease status; and 3) the same reference standard being used to verify all index test outcomes.¹⁶³

Deviation from these criteria results in bias. Colonoscopy with histology is the accepted reference standard for diagnosis of IBD.^{112, 158, 159} The assumption that colonoscopy has sensitivity close to 100%,⁵⁸ however, may be overoptimistic. Although colonoscopy is capable of visualising the terminal ileum, colon, and rectum, a large part of the small bowel remains unseen. It has been estimated that about 30% of patients newly diagnosed with CD may have disease limited to the small bowel beyond the reach of colonoscopy and the disease may also extend to the upper gastrointestinal tract.¹⁵⁸ Using colonoscopy as the reference standard for IBD in FC test accuracy studies may therefore underestimate IBD diagnoses. Nevertheless, the majority of FC test accuracy studies used colonoscopy with or without histology as the reference standard (Table B6). The small number of test accuracy studies of routine data showed that additional investigations were undertaken to confirm disease in clinical practice. I expect that this is the case in my THIN test accuracy study. This means that my study did not rely solely on colonoscopy to confirm disease but was consequently affected by differential verification.

The problem with using routine data, which capture clinical practice, for the reference standard is threefold. Firstly, waiting times for colonoscopy in clinical practice can be months. In this time symptoms may have improved or worsened, meaning that the FC test and reference standard may assess different disease states. Secondly, in clinical practice not all patients with an FC test will receive colonoscopy. Some patients will receive different imaging tests with different sensitivity (leading to bias due to differential verification), and the majority of tested patients, particularly those without a negative FC test, will receive no reference standard (leading to bias due to partial verification). Thirdly, information on secondary care events is not fully captured in the THIN database, which restricted the choice of reference standard for the test accuracy study. In the absence of sufficient information on colonoscopies, I used a proxy for a composite reference standard consisting of various approaches used in clinical practice. My proxy was a combination of Read and drug codes.

de Groot et al. (2011)¹⁶⁴ reported a Bayesian correction method to simultaneously adjust for differential verification bias and for imperfect reference standards. The

model produces test accuracy measures with respect to a latent (theoretically defined) disease status and for separate reference standards. These accuracy measures have greater clinical relevance than a combination of reference standards would achieve. However, the model is developed for prospective studies where the proportions and demographics for the tested negative and positive populations are collected and, therefore, available for modelling. This is not the case for routine data like THIN.

As a result, the reference standard is a major issue for test accuracy studies using routine data. However, a pragmatic study for test accuracy aims to establish the test performance of a test as it is used in clinical practice. The ideal reference standard is generally not feasible in these studies. While this does not invalidate the study, the study aim and the performance of the reference standard needs to be considered in the interpretation of the study findings.

4.3.5.2.4 Confirming disease status in non-IBD cases

The analysis assumed that absence of a Read or drug code for IBD is absence of disease. This assumption created some uncertainty in the test accuracy results as I was unable to test the assumption sufficiently. The proportion of patients without a diagnosis following testing was large and could be due to any of the following: missed IBD; unrecorded IBS; other conditions not considered as study variables; or diagnoses recorded as free text and, thus, not available to the study. Also the research literature was too heterogeneous to be able to estimate the proportion of other conditions accurately. For example, across five studies^{63, 131, 137, 138, 140} the number of conditions varied widely and the proportions ranged from 4% (16/399)⁶³ to 61% (422/694).¹³⁷ I was unable to predict IBD using other variables, such as symptoms and co-morbidities, in the non-IBD group due to the lack of strong predictors of IBD among the study variables. Without any further information on the characteristics of the non-IBD cases, their true disease status remains uncertain and the likelihood of missed IBD could not be asserted any further.

[4.3.5.3 Findings in context of published evidence](#)

The results are broadly in line with pooled estimates from my meta-analysis of 14 secondary and primary care studies (section 4.2.4.1), which may suggest that the pooled test accuracy measures are applicable to the FC tested population in primary care. At the thresholds investigated, my sensitivity and specificity estimates

are within the range of estimates from published primary care studies.⁶¹⁻⁶³ However, previous primary care studies included small numbers of cases while my study provides test accuracy estimates with narrower confidence intervals. Crucially, my study shows that the specificity of 0.94 (95% CI 0.73 to 0.99) reported in the NICE guidance is not applicable to the primary care setting,⁵⁹ where the broader definition of the non-IBD group means that the specificity and the positive predictive value of FC testing is much lower. Therefore, the interpretation of a test positive result may be challenging. A broader disease category in primary care including other organic conditions in addition to IBD resulted in sensitivity estimates ranging from 0.64 to 0.94.^{51, 64, 65} As the conditions may follow different pathways, interpretation of a test positive result would still be challenging. The higher number of patients with a false positive result may not achieve the reduction of referrals to colonoscopy as anticipated by NICE guidance.

4.3.5.4 Suitability of THIN data for test accuracy

The limitations of my THIN test accuracy study may give rise to questions over the suitability of routine data for test accuracy studies. However, as stated in the introduction in section 4.3.1, routine data have become increasingly popular for test evaluations because of their many advantages mainly in overcoming study design issues in situations of low disease prevalence.¹⁵⁰ The key is to recognise the limitations, discuss assumptions, and adjust the claims of the study findings accordingly. Considering the literature^{109, 149} as well as my experience with the THIN database, I believe routine primary care databases such as THIN have merits for certain test accuracy questions, particularly under the following conditions:

1. The index test has a single, specific indication requiring no substantial information on the reasons for testing.
2. The target condition is fatal or serious (not self-limiting), enabling follow-up as a reasonable, alternative reference standard.
3. The index test is established to ensure sufficient data are available yet recent to ensure test results are recorded electronically, which does not favour recording of positive test results.
4. The index test and reference standard are part of primary care and the testing pathway does not cross the interface of primary and secondary care.
5. The target condition is an important, chronic condition which is not deleted during the move of computer systems or of patients between practices.
6. Identification of the target condition from the database has been validated.

In the following I discuss to what extent FC testing for IBD meets these criteria and what this means for the validity of the study findings.

- 1) FC testing this is a good example for a test accuracy study using routine primary care data because it is a very specific test. While the database provides no information on the reason for FC testing, the test should only be ordered for the diagnosis of IBD in a relatively narrow group of people. This is in contrast, for instance, to a white blood cell count which is ordered with every blood test for a range of reasons.
- 2) IBD is a severe condition which, without appropriate treatment, will become more severe and a diagnosis of IBD is likely to be reconsidered within a few months following a false negative FC test.¹⁶¹ This allows follow-up of test negative patients, who do not qualify for the reference standard, without misclassifying missed cases of IBD as true negatives.
- 3) FC testing is not yet fully established in primary care according to my study on uptake (section 3.5.2). However, my study was based on a large number of patients. A future study may look different as GPs are likely to use the test more routinely and adjust who they test. Bias due to preferential recording of test positive results was presumably low (the test positive rate was slightly higher compared to primary care test accuracy studies – see Table 3.11) as test results are now primarily recorded electronically.^{109, 110}
- 4) A major limitation of the FC test accuracy study was the reference standard because secondary care interventions like colonoscopy are not well reported in primary care records. Furthermore, using a recorded IBD diagnosis as a proxy assumes that absence of a recorded diagnosis is absence of disease. Future studies may investigate the negative predictive value of an absent IBD record or consider linkage with additional data to affirm the disease status of tested patients to overcome reference standard issues in the THIN database.
- 5) A chronic condition such as IBD requires ongoing management and treatment. My definition of the target condition included IBD-specific medications. Therefore, my analysis was unlikely to have lost patients who moved or were registered with practices that changed computer systems.
- 6) A study investigating the validity and completeness of using clinical codes to identify cases of IBD reported a positive predictive value of 92%.²⁴ The study used codes from the Oxford Medical Indexing System and a limited number of codes. I did not validate my code list but included a much more detailed list of Read and drug codes.

Considering the limitations and assumptions, the validity of the results of my THIN test accuracy study in isolation is uncertain, mainly due to the issues surrounding the reference standard, excluded tests, and the uncertainty of the true disease status of IBD negative patients. However, the study is a step towards understanding the test performance of FC testing as it is used by GPs in clinical practice. Comparison to my meta-analytical results in section 4.3.4.2.9 showed that point estimates are not unreasonable. However, the meta-analytical result is based on a mix of secondary and primary care studies. Confirmation with meta-analytical estimates applicable to the primary care setting is needed. This is the focus of section 4.4, in which I report the results of my tailored meta-analysis. This involves the identification and inclusion of studies into a meta-analysis, which have been judged to be applicable to the primary care setting.

4.4 Tailored meta-analysis of published test accuracy studies

4.4.1 INTRODUCTION

My meta-analysis in section 4.2 was based on heterogeneous test accuracy studies from primary and secondary care. I considered the applicability of the summary estimates to the UK primary care context to be low. Studies were from different countries and settings and considered different patient populations. In particular, secondary care studies investigated patient populations with a different disease spectrum than would be found in primary care, even if studies investigated referred patients. Firstly, referred patients came from different pathways, including cancer pathways. Secondly, GPs might have tested more patients than they referred. Consequently, the patient populations of the included studies resembled a continuum of primary to secondary care patients. Only some study populations might plausibly resemble the disease spectrum found in UK primary care. In section 4.3 I addressed these applicability concerns by using UK general practice health care records to derive test accuracy measures for FC testing. While applicability for that study could be taken as given, I remain cautious over the reliability of the test accuracy estimates (see section 4.3.5.3). An alternative approach to address these applicability concerns is the tailored meta-analysis.

The tailored meta-analysis^{165, 166} addresses the problem that conventional meta-analyses provide an average of sensitivity and specificity, which may not be applicable to a specific population or setting of interest. The aim of the tailored approach is to identify studies that are truly relevant to the setting of interest based on information collected from the target setting. The information (the test positive rate and disease prevalence) is used to define a plausible range of test accuracy measures for the target setting. This is done by using the mathematical relationship between the test positive rate, disease prevalence, sensitivity, and specificity. Studies compatible with that range are meta-analysed to provide a more relevant pooled estimate for that setting. This has been demonstrated for a national screening programme in the UK context and for a single GP practice.^{165, 166} Tailored meta-analysis of studies evaluating the accuracy of the Pap test for cervical cancer included 12/68 studies from a conventional meta-analysis.¹⁶⁵ The study used the NHS Cervical Screening Programme for information on the test positive rate and prevalence of cervical cancer in the UK.¹⁶⁵ From conventional meta-analysis, sensitivity and specificity for Pap testing for cervical cancer were 72.8% (95% CI 65.8% to 78.8%) and 75.4% (95% CI 68.1% to 81.5%), compared with 50.9% (95%

CI 35.8% to 66.0%) and 98.0% (95% CI 95.4% to 99.1%) from a tailored meta-analysis. The tailored result is likely to be more plausible in the NHS.¹⁶⁵ Tailored meta-analysis has also been shown to provide results more compatible with the population of a single GP practice by collecting practice specific data.¹⁶⁶ Five out of 12 non-UK primary care studies evaluating the accuracy of Centor's criteria for streptococcus infection in sore throats were deemed plausible for the GP practice. Tailored meta-analysis of the five studies resulted in sensitivity and specificity of 38.4% (95% CI 30.3% to 47.2%) and 92.1% (95% CI 83.4% to 96.5%). In contrast, the sensitivity and specificity estimated from the conventional meta-analysis were 50.3% (95% CI 42.7% to 58.4%) and 78.5% (95% CI 65.7% to 87.5%). This illustrates that conducting studies in a similar setting does not guarantee that findings will be applicable.

I used the tailored meta-analysis approach to select studies that were plausible for UK primary care irrespective of their reported setting. I based the tailoring on setting specific information from routine primary care. Therefore, my third approach to estimate sensitivity and specificity of FC testing for IBD in primary care combined published evidence from the systematic review (section 4.2) with information from my THIN study (section 4.3). My aim was to gain some insight into the direction of bias in conventional meta-analyses when spectrum bias is suspected. In particular, I aimed to test the hypothesis that meta-analysed estimates from selected studies will move closer to the THIN derived estimate of test accuracy and, with that, be able to validate the tailored meta-analysis approach for the first time.

4.4.2 OBJECTIVES

- To define the applicable region (the plausible range of sensitivity and 1-specificity) in ROC space for FC testing for primary care
- To select studies from the systematic review that are applicable for primary care
- To derive tailored estimates of sensitivity and specificity for FC testing for primary care

4.4.3 METHODS

4.4.3.1 Studies considered for tailored meta-analysis

The included studies from the systematic review in section 4.2 formed the basis for the tailored meta-analysis. There were 14 studies reporting test accuracy measures

at the 50µg/g threshold for the differentiation of IBD versus non-IBD (section 4.2.4.1). I plotted the sensitivity and 1-specificity pairs of the 14 studies to visualise their position in the ROC space and in relation to the applicable region.

I considered new studies from auto alerts up until March 2019.

4.4.3.2 Definition of test positive rate and IBD prevalence

I used the dataset of 7,084 FC tested patients described in section 4.3.4.2 to calculate the test positive rate and IBD prevalence in FC tested patients in primary care. I defined the test positive rate as the proportion of FC tests with a numeric FC value of >50µg/g. Prevalence was defined as the proportion of IBD diagnoses in the FC tested population. 99.98% confidence intervals for test positive rate and prevalence were calculated using the Hotelling method.¹⁶⁷ This method takes into consideration the correlation between the prevalence and test positive rate. Using the Hotelling method ensured that the coverage of the 99.98% CI was at least 99.98%.¹⁶⁷ As the confidence intervals were used to define the applicable region, this was the most conservative method to maximise the chance for studies to be compatible with the applicable region and be included into the tailored meta-analysis.

4.4.3.3 Selection of studies for tailored meta-analysis

The selection of the studies required two steps. Firstly, I used the information on the FC test positive rate and the IBD prevalence collected from the THIN dataset to define an applicable region in the ROC space. This is described in detail in Willis et al. (2014 & 2015)^{165, 166} and summarised in Appendix B15. The applicable region represents the area of a plausible range of sensitivity and 1-specificity (false positive rate) pairs for FC testing in primary care. Secondly, I compared the studies' test accuracy estimates with the applicable region for each study. Studies with test accuracy estimates lying within this region were considered applicable for the primary care setting. For studies with estimates lying outside the applicable region, I determined the probability that their population parameters were within the applicable region (see Figure B8 in Appendix B15). This assessment was based on maximum likelihood estimation described by Willis et al. (2014).¹⁶⁵ In cases where the probability was smaller than 0.025 the study was rejected.

I defined the applicability region and selected studies for meta-analysis in R using the R code provided by Dr B. H. Willis (see Appendix B16).

4.4.3.4 Tailored meta-analysis

The meta-analysis considered test accuracy studies of FC testing for IBD at a 50µg/g threshold to derive summary estimates of sensitivity and specificity. I used R version 3.6.1 to undertake a bivariate random-effects meta-analysis¹⁰⁴ including only studies that were plausible to lie within the applicable region, i.e. that were compatible with the test positive rate and prevalence found in primary care. I compared tailored estimates with those from my conventional meta-analysis and my THIN test accuracy study.

4.4.4 RESULTS

4.4.4.1 Selection of studies for tailored meta-analysis

In my systematic review in section 4.2 I identified 14 studies evaluating FC for the differentiation of IBD and non-IBD at the FC threshold of 50µg/g.^{33, 61-63, 130, 131, 134, 137, 138, 140, 142-144, 146} I considered all studies for the tailored meta-analysis. I identified four more recent studies from primary care from auto alerts.^{64, 65, 168, 169} However, they reported results for a different clinical question and/or a different threshold and were not considered in the tailored meta-analysis.

Figure 4.18 shows the 14 studies in the ROC space and in relation to the applicable region. The applicable region marks the area of greatest plausibility for primary care based on THIN data. The applicable region was narrow due to the precision introduced by using the large THIN dataset. None of the study estimates actually lay in the applicable region. Half of the studies had a low probability of their study parameter to lie in the applicable region.^{33, 61, 62, 131, 134, 137, 146} These studies were excluded from meta-analysis because they were outside the range of performances possible for primary care. The remaining studies reported test sensitivity and false positive rate that were compatible with the applicable region and were included.^{63, 130, 138, 140, 142-144} Only one of three studies from primary care (Conroy 2018⁶³) was included. The study reported the lowest sensitivity of FC testing for IBD compared to the remaining 13 studies.

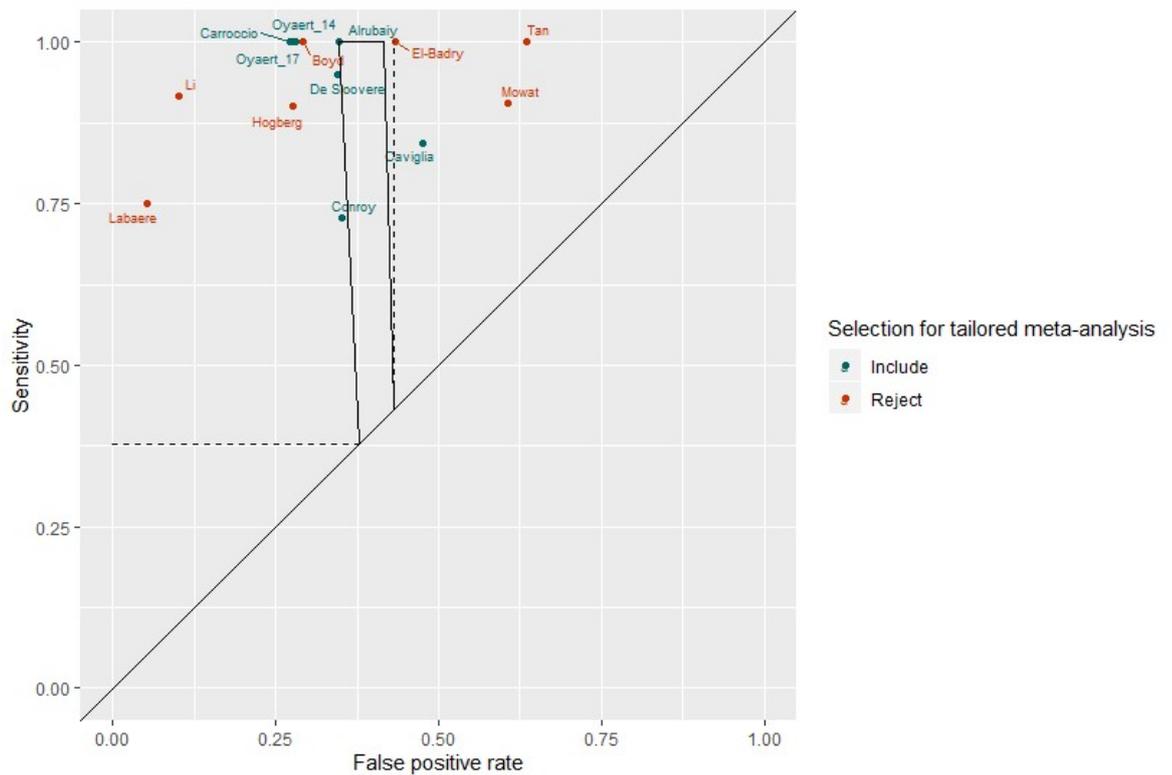


Figure 4.18 ROC plot of studies reporting the sensitivity and specificity of FC testing for IBD at the threshold of 50 μ g/g. The applicability region for primary care defined by the test positive rate (dashed line) and by test positive rate plus prevalence (trapezium) from THIN data defines the area of sensitivity and specificity most plausible for THIN practices.

4.4.4.2 Tailored meta-analysis

The results of the tailored meta-analysis of seven studies in comparison to the results from my conventional meta-analysis are shown in Table 4.11. Sensitivity and specificity from tailored meta-analysis were 0.97 (95% CI 0.81 to 0.996) and 0.68 (95% CI 0.64 to 0.72), respectively. The corresponding estimates from my conventional meta-analysis were 0.94 (95% CI 0.90 to 0.97) and 0.67 (95% CI 0.57 to 0.76). The point estimates of the conventional meta-analysis were closer to the estimates using THIN data than the tailored meta-analysis results. However, confidence intervals overlapped (Table 4.11). Figure 4.19 reveals that the THIN estimate lay in the applicability region. However, both results from meta-analysis lay outside the applicability region. This illustrates the possibility that tailored meta-analysis does not always produce more applicable results than conventional meta-analysis. This is most likely due to the majority of studies included being small (wide confidence intervals) and lying on the left hand side of the narrow applicability region.

Table 4.11 Comparison of sensitivity and specificity at the common threshold of 50µg/g from tailored meta-analysis with conventional meta-analysis and results using THIN data

Outcome	THIN data (95% CI) (section 4.3.4.2.2)	Conventional MA (14 studies) (95% CI) (section 4.2.4.1)	Tailored MA (7 studies) (95% CI) (section 4.4.4.2)
Sensitivity (95% CI)	0.93 (0.89 to 0.96)	0.94 (0.90 to 0.97)	0.97 (0.81 to 0.996)
Specificity (95% CI)	0.61 (0.6 to 0.63)	0.67 (0.57 to 0.76)	0.68 (0.64 to 0.72)

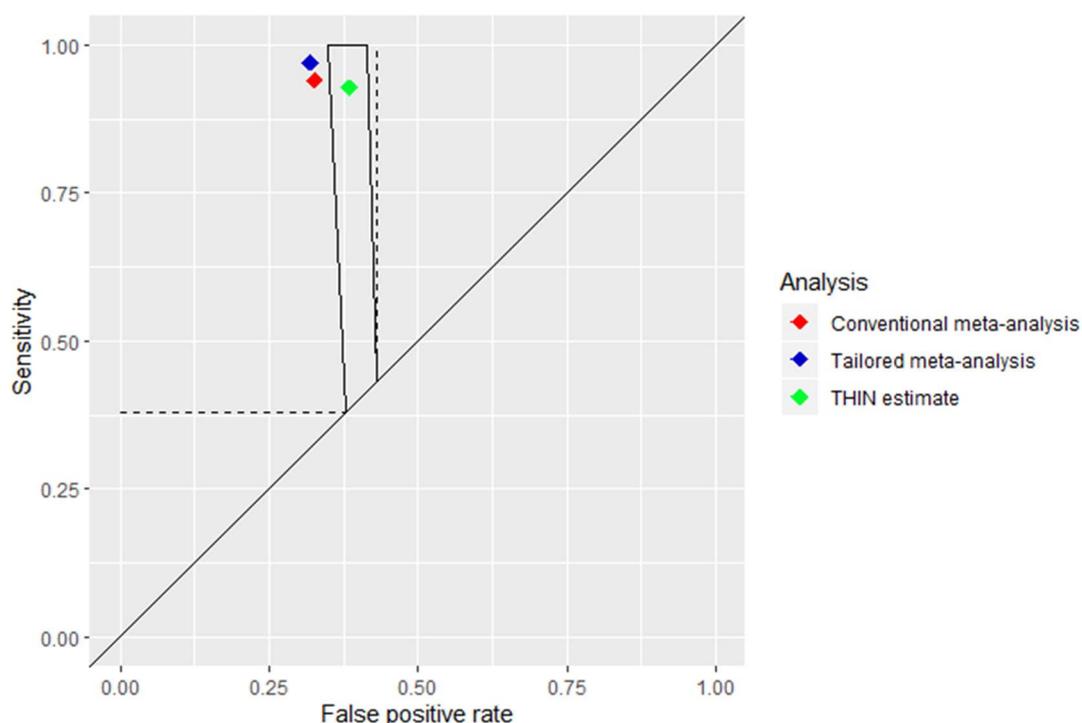


Figure 4.19 Pairs of sensitivity and false positive rate from conventional meta-analysis, tailored meta-analysis and THIN data in ROC space and in relation to the applicable region informed by routine data from THIN

4.4.5 DISCUSSION

4.4.5.1 Summary of study findings

Seven of 14 test accuracy studies identified for conventional meta-analysis were deemed to be applicable to the primary care setting. None of the included studies lay in the applicable region and only one of three primary care studies were included in the tailored meta-analysis. This demonstrates that identifying a study as a primary care test accuracy study does not guarantee that study outcomes are actually applicable to the setting of primary care defined by THIN data. The tailored meta-analysis of the seven studies resulted in estimates of sensitivity and specificity of 0.97 (95% CI 0.81 to 0.996) and 0.68 (95% CI 0.64 to 0.72). The estimates were

further away from the THIN estimate of 0.93 (95% CI 0.89 to 0.96) and 0.61 (95% CI 0.6 to 0.63) than the estimates from conventional meta-analysis including all 14 studies. Furthermore, the tailored average did not lie in the applicable region. My findings presented an example where the tailored result was not more plausible for the target setting than the result from conventional meta-analysis. However, estimates were close and confidence intervals overlapped.

4.4.5.2 Limitations of the study

In my study the tailored result was not closer to the THIN estimate than the conventional result. The finding is not what I had anticipated, as the tailored result is based on more information than the conventional meta-analysis. This reveals a possible weakness of the method of tailored meta-analysis, which currently excludes studies deemed implausible. The method should be adjusted to account for the situation where the tailored result lies outside of the applicable region. Firstly, all available studies should be included and weighted according to the probability that their estimates are plausible. Secondly, the constraints that define the applicable region could be incorporated into the bivariate random effects model (rather than using them to select studies before meta-analysis). This would ensure that the meta-analysis estimates are always in the applicable region.

The success of the method relies on the fact that the applicable region is correct. This requires accurate estimates of the test positive rate and IBD prevalence. In section 3.5.3.2 I reported that the test positive rate using the THIN data is slightly greater compared to those reported in primary care FC test accuracy studies. Furthermore, in section 4.3.4.1 I showed that there is a great proportion of FC tests with missing results. It may be possible that positive test results are recorded with more diligence than test negative results in general practice which in turn would result in higher test positive rates. This creates some uncertainty about the estimate of the test positive rate used in the tailored meta-analysis.

The IBD prevalence relies on accurate and complete coding of IBD in primary care records. As discussed in section 3.6.2.3 potentially missing codes could not be identified or quantified which casts a shadow of doubt on the credibility of the prevalence of IBD used to define the applicable region. However, IBD prevalence in FC tested patients was similar to that in studies of FC tested patients (Table 3.11).

Uncertainty in estimates of test positive rate and prevalence may have led to incorrect boundaries being drawn for the applicable region. However, this was mitigated, as suggested by Willis and Hyde (2014),¹⁶⁶ by using 99.98% CI intervals with high coverage probability to maximise the probability of studies being included.

4.4.5.3 Findings in context of published evidence

The estimates from tailored meta-analysis and conventional meta-analysis were similar considering their confidence intervals. This may be due to the fact that the applicable region excluded studies on both sides of its boundaries. Since both estimates represent an average of the included studies excluding studies from both sides did not have a significant effect on the averages. This is in contrast to tailored meta-analyses in the literature, where studies outside the applicable region were either all outside the left boundary or all outside the right boundary of the applicable region.^{165, 166} Furthermore, examples in the literature all had some studies lying inside the applicable region. This was true for narrow applicable regions informed by routine data from UK screening programmes as well as wider applicable regions informed by limited UK data from a single general practice. In these previous examples, estimates from tailored meta-analysis lay within the applicable region and presented better estimates for the setting of interest. However, none of the published studies compared the tailored and conventional result with an estimate from the setting itself because data for the disease status following testing were not available (or collected).

Conventional meta-analyses are routinely presented with a confidence ellipse and a prediction region around the estimate of test accuracy in ROC space.¹⁷⁰ The prediction region is the region where a new study's estimate would lie. This is different to the applicable region because the prediction region is based only on information from the meta-analysed studies while the applicable region utilises data from the actual setting of interest and is, therefore, more specific.

My conventional meta-analysis was affected by the heterogeneity in patient mix across studies. However, it was not possible to measure its effect on the test accuracy measures. Tailoring the inclusion of studies decreased the heterogeneity. This is a clear advantage of tailored meta-analysis. However, tailoring did not result in the expected difference in estimates from tailored and conventional meta-analysis. I was, therefore, unable to estimate the direction and extent of spectrum bias we may introduce by accepting secondary care evidence for the evaluation of

test accuracy in primary care. Furthermore, I cannot claim that the tailored result is more accurate because it did not lie in the applicable region. Therefore, my study could not validate the tailored meta-analysis as an approach to produce test accuracy measures that are closer to estimates from the target setting.

4.5 Summarising the test accuracy of FC testing

In this chapter I investigated the test accuracy of FC testing for IBD in three studies using three different methods. In a systematic review (section 4.2) I showed that 1) test accuracy of FC testing was high irrespective of assay type or clinical question; 2) sensitivity of FC testing in the differential diagnosis of IBD versus non-IBD was high but specificity was modest; and 3) the sensitivity in primary care may be lower than in secondary care based on results from individual studies in the absence of a meaningful comparison of meta-analysed results. Overall, I assessed the applicability of included studies to the primary care context to be low.

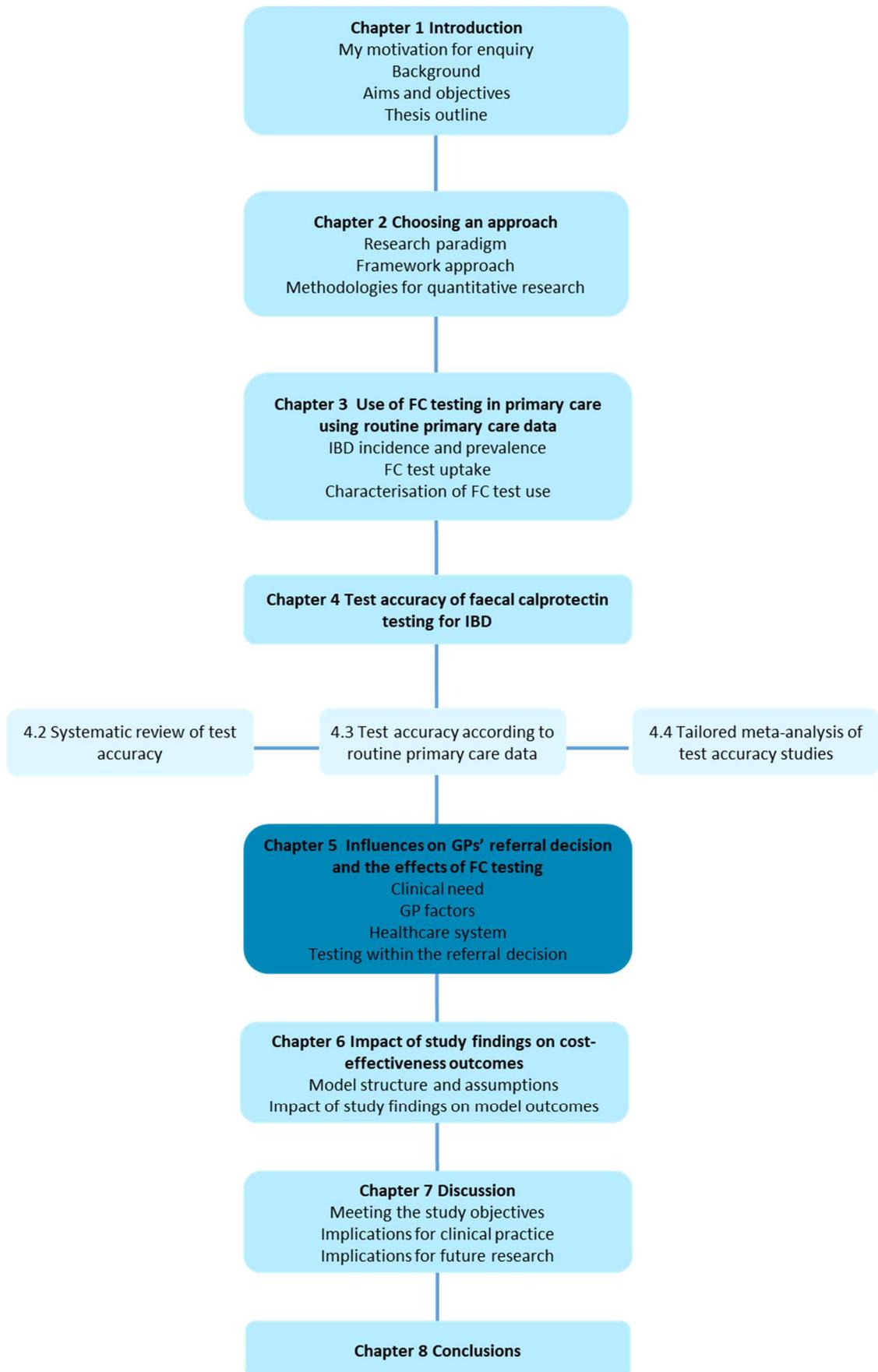
In my test accuracy study of routine data on FC testing for IBD in primary care (section 4.3) I reported high estimates for sensitivity but modest estimates for specificity. The estimates were within the range of estimates from primary care studies and close to the meta-analysed result of studies from primary and secondary care. However, a considerable proportion of FC tested patients with missing test results were excluded and I was unable to confirm the absence of IBD in cases without IBD diagnoses. The magnitude and direction of bias of the results of the THIN test accuracy study are unclear.

Finally, I combined published evidence from the systematic review (section 4.2) with setting-specific information from routine data (section 4.3) in a tailored meta-analysis (section 4.4). In this analysis, sensitivity of FC testing was again high and specificity was modest. The tailored meta-analysis can be used as an extra source of information when triangulating results to enable conclusions about test accuracy. Including additional setting-specific information to tailor study inclusion in the meta-analysis did not, as anticipated, result in test accuracy estimates that were closer to the THIN estimate. This may be because the method excludes studies with a low probability of being compatible with the primary care setting, rather than weighing them by their probability.

In all three studies, the sensitivity of FC testing was high (Table 4.16). Due to the high number of false positives, specificity was modest. The point estimates of the THIN analysis were lowest and the point estimates for the tailored meta-analysis highest. However, confidence intervals overlapped, which precludes the conclusion that estimates were truly different. Therefore, I did not find conclusive evidence to support my hypothesis that estimates from meta-analysis including studies from secondary care, primary care and of referred patients are not applicable to the primary care setting.

The positive predictive value of FC testing was low in primary care due to the high number of false positive test results. The group of patients with false positive results comprise patients with no diagnosis, IBS or other inflammatory conditions because FC testing detects inflammation rather than specifically IBD. Interpretation of a positive test result may, therefore, differ in a test accuracy study and in general practice. The negative predictive value was very high. In a low prevalence setting like primary care, the few false negative cases are masked by the great number of true negative cases. Increasing the threshold to 100µg/g created an additional 7% (14/210 IBD cases) of false negatives. We need to understand the impact of testing on patient management decisions in order to judge whether the additional missed IBD cases are acceptable.

The high negative predictive value was incompatible with the frequent referral of test negative patients as seen in section 3.5.4.2 and section 4.3.4.2.8. My assessment of referral by disease classification status following testing suggests that GPs use other cues in addition to FC testing in management decisions. We need to understand the GPs' reasons for referral. Furthermore, the high proportion of patients tested with a previous IBS diagnosis (Figure 4.12) revealed the need to understand the GPs' reasons for testing. The studies of test accuracy do not support any inferences about why GPs test and refer. This requires a different approach using qualitative methods which I describe next in Chapter 5.



Chapter 5 Influences on GPs' referral decision and the effects of FC testing

5.1 Chapter overview

In Chapters 3 and 4, I investigated how FC testing is used in clinical practice and how accurate it may be in primary care. The studies did not provide information on why GPs use FC testing, when they consider it useful, and how FC testing influences the referral decision. However, given the apparently inappropriate referrals of FC test negative patients described in Chapter 3, it seems imperative to understand how GPs make a referral decision and what influences their decision to refer. This requires a different methodological approach. In this chapter I describe my interview study with 19 GPs to explore how GPs make referral decisions and how FC testing may fit into the decision. I start this chapter with the rationale for the study followed by some background on referral and referral decision models. This is followed by the methods explaining the framework approach,⁷² how GPs were recruited to ensure a diverse sample and how the patient perspective was captured. The results are organised by theme, first covering the identified influences on the referral decision followed by the outcomes of FC testing. In the discussion, I first consider the main findings in the context of published literature. I then highlight the strengths and limitations of the study and end with a few thoughts on validating my qualitative research and a personal reflection.

5.2 Introduction

5.2.1 RATIONALE

The development of decision support tools such as tests and health economic models require evidence-based theories that explain the underlying mechanisms of human judgment and decision making.¹⁷¹ This is important to allow generalisation of such tools to a range of health care decisions and to different settings. Such theory, however, is lacking in research on medical decision making, instead the development of decision support tools is often based on assumptions about human judgment and decision making, therefore, limiting their use for real time decision making.¹⁷¹ This is evident in the current evaluation of FC testing. To date, it is largely unknown how the FC test is used and how it impacts GPs' clinical decision making. The NICE cost effectiveness analysis assumed that GPs only refer patients with a positive FC test to secondary care.⁵⁸ This was based on the expectation that GPs' referral behaviour would follow research on test accuracy, costs and benefits.

However, more recent studies reporting experiences with FC testing in primary care^{55, 63, 64} and my findings in section 3.5.4.2 and section 4.3.4.2.8 suggest that some GPs do not act on FC test outcomes and do not follow this idealised behaviour. An understanding of the mechanisms of human judgement and decision making to explain GPs' actual behaviour is key in making research on decision tools valid and generalisable.¹⁷¹

5.2.2 BACKGROUND

About one in 20 GP consultations results in referral.¹⁷² Research showed that there is a wide range in patients referred per practice (2.3-24.4%) and most of this variation is still unexplained.^{173, 174} Variation could not be sufficiently explained by characteristics of the patient (age, sex, social class, case mix), characteristics of the GPs (age, experience, background, specialist interest, workload) or characteristics of the practices (size, ancillary staff, fundholding, proximity to hospital).^{174, 175} More of the unexplained variation occurs within than between practices.¹⁷³ This suggests that the GPs' decision of whether to refer a patient varies. Variation in referral rates has been the topic of investigation since the early 1960s because of concerns about cost implications and potential negative effects on patients.^{176, 177} Newton et al. (1991)¹⁷⁶ argued that the reason for the lack of understanding is the positivist view in the research undertaken to address the issue and that referral decisions need to be understood from the GPs' perspective and that of their patients. The two initial studies that aimed to understand the referral decision-making process concentrated on referral for diagnosis¹⁷⁸ and referral for diagnosis, treatment or advice.¹⁷⁵ However, the decision making process is even more complex. Newton et al. (1991) uncovered that GPs may refer patients where referral does not appear necessary on clinical grounds to "share the burden of care".¹⁷⁶ They concluded that we are far from being able to explain referral patterns. And even though ongoing research has tried to understand variation in referral rates using routine data,^{111, 121, 173} surveys,¹⁷⁹⁻¹⁸¹ and qualitative research,¹⁸²⁻¹⁸⁴ I have been unable to identify a study specifically looking at referrals for IBS and IBD patients. There is, therefore, a need to investigate the reasons for referral of patients with IBS or IBD and to understand the views and experiences of GPs related to referral and FC testing. Such insights can help explain the apparent inappropriate referral of FC negative patients.

5.2.3 THEORETICAL FRAMEWORK

Medical decision making and problem solving have been studied for many years. Early models have shed some light on the cognitive aspects involved in clinical

reasoning and on how practitioners process information.¹⁸⁵⁻¹⁸⁷ However, the models focused on diagnostic decisions in acute settings, and it has been questioned whether the theories are applicable to other types of medical decisions.¹⁸⁸ The model by Dowie (1983)¹⁷⁸ was the first attempt to explain referral decision making under conditions of uncertainty. Dowie described the referral decision as a process. Her model combined medical judgment and sociological variables in a series of questions. Dowie's model has been criticised for its focus on acute patients and the decision to refer to obtain a diagnosis rather than referring for treatment or advice.¹⁷⁵ The process of making a referral and the implications will vary according to these different reasons of referral.¹⁷³ For Newton et al. (1991)¹⁷⁶ referral is a social (inter-)action, rather than a process. King et al. (1994) also questioned whether decision making really follows such a rational approach as is suggested by the process models.¹⁸⁹

The objective of this study was to understand the individual referral decision making process from the GPs' perspective. In particular, I wanted to explore GPs' reasons for referral and influences on referral decisions in the context of individual experiences with IBS and IBD patients. While there is existing literature on clinical decision making, problem solving and referral decisions, in my study I did not aim to deductively prove or disprove existing models or hypotheses derived from this literature. This study adopted an inductive approach to understand the views and experiences of GPs in relation to referral decisions of patients with either IBD or IBS in order to explore the role of FC testing in the decision making process. The study aimed to add to our knowledge of GP referral decisions in the specific context of IBS and IBD and contribute an understanding of how a specific test affects referral decisions.

5.3 Objectives

- To understand how GPs make referral decisions
- To explore how testing fits into the referral decision
- To explain the reasons for the difference between actual and idealised referral behaviour which questions the validity of the cost-effectiveness estimate of FC testing in primary care

5.4 Methods

5.4.1 STUDY DESIGN

Using framework methodology,^{72, 78} I conducted semi-structured interviews with 19 GPs to explore their views and experiences related to referral of patients with chronic abdominal symptoms.

5.4.2 ETHICAL CONSIDERATIONS

I obtained study approval from the Health Research Authority and the University's Biomedical and Scientific Research Ethics Committee (REGO-2017-1914).

5.4.3 SAMPLE AND RECRUITMENT

5.4.3.1 Eligibility Criteria

I undertook interviews with GPs practising in clinics within the West Midlands South region (as defined by the Primary Care Clinical Research Network [CRN]) consisting of seven CCGs covering Coventry, Warwickshire, Herefordshire and Worcestershire. GPs were from practices with FC testing available and with no FC testing available. Sampling aimed for a diverse sample (maximum variation sampling) taking into account a range of experience (full time equivalent of years in practice), ethnicity, gender, size of GP practice, location of GP practice (urban versus rural) and previous experience with FC testing. The aim was not to have representation of all possible combinations of characteristics but to have a diverse sample.

Inclusion criteria

- Gender: both male and female GPs
- Range of experience as a practising GP (years full time equivalent)
- Ethnicity: any
- Location: urban, rural, mixed, different sized practices
- Experience with FC testing (ranging from none to frequent user)

5.4.3.2 Sampling

My aim was to conduct interviews with about 20 GPs.^{190, 191} This decision was based on Mason (2010)¹⁹⁰ and Guest et al. (2013)¹⁹¹ who argued that interview studies should not include less than 15 interviewees. Yet, interviewing more than 20 will usually not generate evidence that will drastically change conclusions.^{190, 191} After about 14 interviews I felt that no new concepts were emerging from the

interviews and that I was reaching saturation. I completed the remaining interviews that had already been scheduled. This allowed me to explore the concept of a wasted referral further and to establish more clarity around urgent and routine referrals. Following that, I decided that no more interviews were required.

5.4.3.3 Recruitment

Identification and recruitment of participating GPs was facilitated by the Primary Care CRN West Midlands South. Five GP research champions (GPs who have agreed to work closely with the CRN to promote primary care research) from five West Midlands South CCGs were recruited via the CRN and sent the letter of invitation and the participant information leaflet. The CRN further facilitated recruitment by e-mailing GPs in the region. Recruitment was an iterative process. I collected demographic data and practice characteristics prior to the interview. Once an initial sample was established, I used the information collected for more targeted recruitment aiming for a diverse sample. Participating GPs were offered a payment of £80.00/interview for the interview to enable backfill of their time.

5.4.4 DATA COLLECTION

I interviewed GPs once face to face in their practices (one interview took place in a CCG office and one at my University). The interviews lasted between 40 and 70 minutes and were electronically recorded with a voice recorder. My open questions aimed to initiate a conversation in which GPs were invited to tell their story from their perspective with a few prompts to encourage more detailed responses.

The interview schedule (Appendix C1) covered questions on the following topics:

- Experience with / perception of IBD/IBS patients
- GPs' perceived role in diagnosing IBD/IBS
- Reasons for referral
- Influences on and thinking behind referral decisions
- GPs' perception of uncertainty and risk
- Role of the patient in referral decisions
- Importance of test results, such as the FC test, in the referral decision

I designed and refined questions for each topic with the input of my Patient and Public Involvement (PPI) advisors and with advisory group members with qualitative research expertise or GP background. As a result I started the interviews with

questions focusing on IBS (the more common clinical presentation), added a question on the perceived impact of symptoms on the patients' quality of life and added a question on why/why not GPs use FC testing. I piloted the interview questions with a GP working at Warwick University before finalising the interview schedule.

5.4.5 ANALYSIS

I transferred the interviews onto a password protected computer and all data were transcribed verbatim. I transcribed the first three interviews myself while the remaining were transcribed professionally.

I undertook the following stages of the framework analysis approach:

5.4.5.1 Familiarisation with the data

I repeatedly read and listened to the interviews and checked the professional transcripts against the recorded interviews.

5.4.5.2 Coding

I started coding as soon as I had completed the first interview using open coding line by line for the first five transcripts.

5.4.5.3 Developing a working analytical framework

I used the initial codes to develop a set of codes (axial coding) organised into categories which were informed by the data. I added new codes to the framework throughout the time I conducted the interviews. I undertook an exercise with the PPI advisors in which I asked them to categorise a subset of line by line codes in order to discuss and validate the developing framework. I used the messages from this exercise to finalise the framework. The final framework is available in Appendix C2.

5.4.5.4 Applying the analytical framework

I indexed (coded) the subsequent transcripts in NVivo using the existing categories and codes of the analytical framework. I added four more codes to the analytical framework during indexing.

5.4.5.5 Charting data into the framework matrix

Once I had coded all transcripts, I developed matrices for each category in Excel with codes as column headings and interviewees as rows. I charted and summarised the coded information from each transcript into the spreadsheets. I added *in vivo* codes (GPs' actual words) and initial analytical thoughts.

5.4.5.6 Interpreting the data

I considered each column of my framework matrices in turn. I identified all elements within each column, ordered the information by emerging analytical concepts and removed duplicate ideas. Emerging analytical concepts developed into key dimensions which I grouped into categories and sub-themes. I used the constant comparison method to validate my increasingly abstract ideas. During this process, themes emerged from the data which could house the subthemes. I further refined themes and subthemes by collapsing, splitting and moving columns. The final categorisation of coded elements into increasingly higher order categories and subthemes are available for the "doctor-patient interaction" theme in Appendix C3. This analytical table illustrates my journey from codes to sub-themes and to themes. I used these tables to ask my PPI advisors for feedback on my analysis and to add any information missing from the patients' perspective. These comments did not change the analysis but determined the direction of some aspects in the discussion of the findings. During the process, I used analytic memo writing to explore emerging ideas. I used the memos and analytical tables to write the final text summarising my interpretation of the data in which I included direct quotes to allow readers to evaluate the inferences drawn from the data. A summary of the preliminary findings was sent to the 19 GPs for member check.

5.4.6 PATIENT & PUBLIC INVOLVEMENT

I recruited PPI advisors by contacting the national charities for IBS (IBS network) and IBD (Crohn's and Colitis UK) who posted my advert on their website (IBS network) or sent out my invitation to participate to an existing group of PPI representatives (Crohn's and Colitis UK). Eight patients were actively involved during the process of the funding application to the NIHR. Four patients attended meetings throughout the project and an additional four interested patients, who were unable to attend meetings, received regular updates on the project as well as invites to comment on documents as required. The aim was to get the patients' perspective on the research to ensure the focus was appropriate, acceptable and relevant to

them, to identify issues of importance to patients, to contextualise the study and the study findings.

Patients were actively involved in the management and dissemination of the research in what is described by INVOLVE as collaboration.¹⁹² Patients were invited to:

- Contribute to discussions at PPI meetings
- Contribute to the development of GP interview questions
- Assist with interpretation of study findings considering their knowledge and experience
- Comment on documents summarising study findings
- Advise on the dissemination of study findings
- Contribute to the writing of articles for patients and the public

5.5 Results

Between August 2017 and January 2018 I interviewed 19 GPs from the West Midlands South region which is illustrated in Figure 5.1. Table 5.1 and Table 5.2 summarise the range of the characteristics of the participating practices and GPs including GPs with and without experience with FC testing.

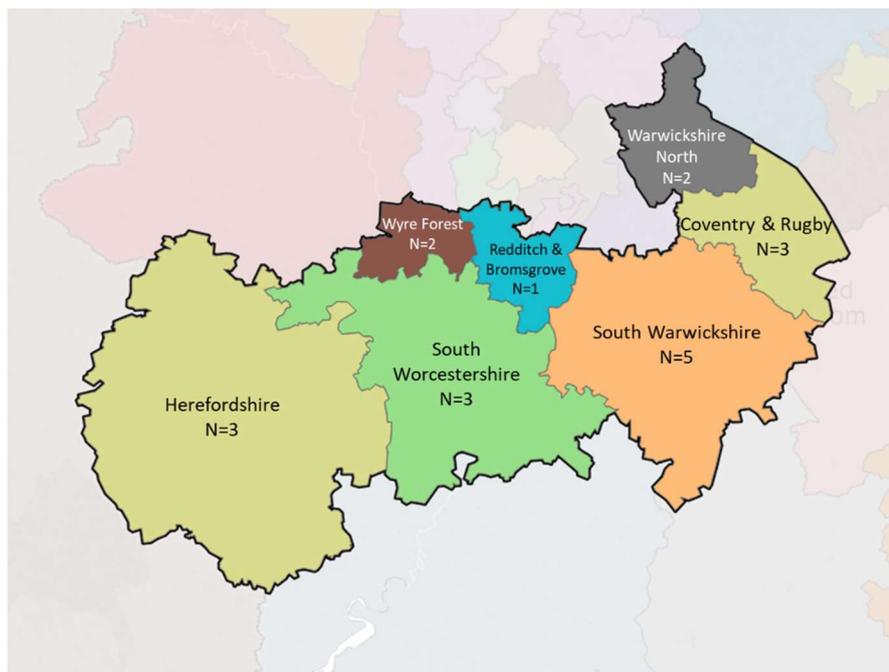


Figure 5.1 Map of recruitment from the seven Clinical Commissioning Groups (CCGs) of the West Midlands South region; depicting the number of GPs recruited from each CCG

Table 5.1 Characteristics of participating general practices

Practice Characteristic	Number of practices in interview
List size (number of patients registered with general practice)	
<5,000	3
5,000 to ≤10,000	9
10,000 to ≤15,000	5
>15,000	2
Deprivation of catchment area	
1 most deprived	0
2	1
3	0
4	2
5	5
6	3
7	2
8	4
9	1
10 least deprived	1
Availability of FC testing	
Available	13
Unavailable	6
Location	
Urban	9
Rural	4
Mixed	6

Table 5.2 Characteristics of participating GPs

GP Characteristic	Number of GPs in interview
Gender	
Female	7
Male	12
Age	
21-30	1
31-40	6
41-50	8
51-60	3
>60	1
Ethnicity	
White	13
Asian	4
Black	0
Other	2
Experience (years FTE)	
≤5 years	2
>5 to ≤10 years	7
>10 to ≤20 years	6
>20 years	4
Speciality	
Gastroenterology	2
General Practice	14
Other	3
Association with practice	
Partner	16
Salaried	3
Experience with FC testing	
I have not heard of FC testing	1
I have heard of FC testing but have not used it	7
I sometimes use FC testing	10
I frequently use FC testing	1

FC faecal calprotectin, FTE full time equivalent, GP general practitioner

Four major themes emerged from the data which describe the main influences on a referral decision (Table 5.3). These four influences are clinical need, the doctor-patient interaction, GP factors and the health care system. The analysis revealed

that referral decisions are not linear. They are complex because the influences are interdependent and while all four influences operate on every decision they can carry different weight in individual situations.

Table 5.3 Overview of themes and subthemes describing the influences on the referral decision

Themes	Subthemes
Clinical need	Unambiguous clinical need (under clinical certainty) Ambiguous clinical need (under clinical uncertainty)
Doctor-patient interaction	GP led Balance of interests Patient pressure Societal pressure
GP factors	GP's mind set GP's experience GP's motivation
Health care system	Direct influence (through active measures) Indirect influence (through mode of service delivery)

Referral decisions can be made with and without testing. If a test is indicated the result provides additional information to inform the management decision which may involve the referral of the patient. A test result is not considered in isolation, however. It is regarded within a wealth of information from history and examinations. The weight of a test result in a management decision depends on a multitude of factors including the reason for testing, the test outcome, prior belief, available guidelines and patient expectations to name a few. In my overview of referral influences (Figure 5.2), I, therefore, do not consider FC testing as an additional influence on the referral decision as on the same level as the four influences clinical need, doctor-patient interaction, GP factors and the health care system. Rather, the various outcomes of testing, both positive and negative, link into aspects of the four influences and may indirectly be supportive or unsupportive of the referral decision or have no impact on the referral decision at all.

I build the case for this interpretation in the following section which describes the four referral themes and the outcomes of FC testing that link testing into the referral decision. I finish with a discussion of the role and place of FC testing in primary care and GPs' requirements for testing.

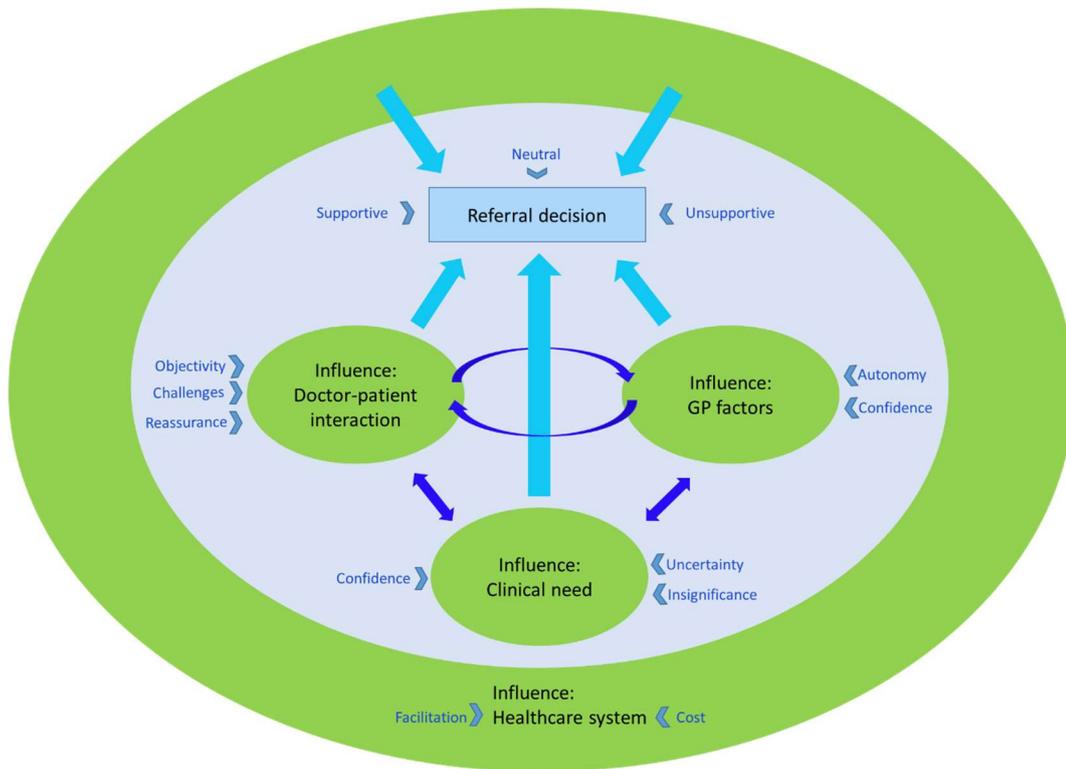


Figure 5.2 Overview of four influences (green ovals) on referral decisions and outcomes of FC testing (◀)

5.5.1 INFLUENCES ON REFERRAL DECISIONS

5.5.1.1 Clinical need

“We’re referring, (...) either because we think there’s something urgently needs dealing with or the ones that we don’t know what’s going on.” (GP14)

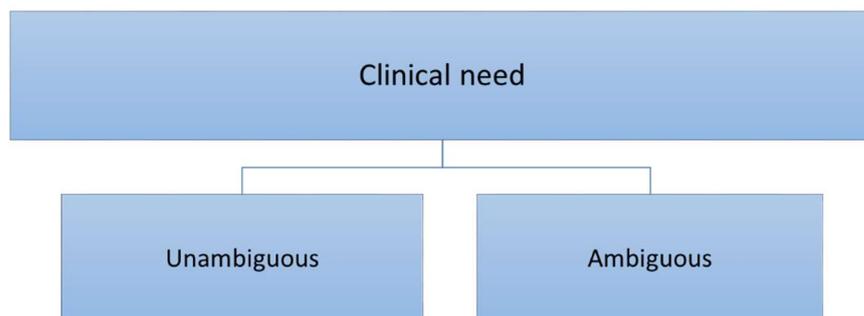


Figure 5.3 Theme 1 of referral influences - clinical need - and its subthemes

My analysis of the interview data suggested that the decision to refer is determined by clinical need as defined by the GP and influenced by the views of patients and secondary care consultants. Clinical need for a referral, in my data, was established

through history, examination and investigations by considering the severity of symptoms, signs, frequency of attendance and more subtle cues and patients' mannerism. Clinical need was seen as either **unambiguous** when in response to either clinical or diagnostic certainty, or **ambiguous** when clinical uncertainty caused apparently inappropriate or intuitive referrals (Figure 5.3).

5.5.1.1.1 Unambiguous clinical need

My interview findings suggested that the GPs' main priority is to manage patients appropriately. Therefore, diagnosis was frequently discussed as a means to manage the patient. In order to arrive at a diagnosis, GPs described putting serious diseases at the top of their diagnostic pyramid and exclude them in turn using different approaches of clinical reasoning. In my data, diagnostic certainty came with clear signs and symptoms or other cues like new presentations of acute symptoms. This certainty can then lead on to clear-cut referral decisions.

*"I always, in my head, have a pyramid of pathology. So, in my pyramid, at the top of the pyramid, they're the nasty things, cancers, things like that and then I have to rule that out (...), it's deadly illness until proven otherwise."
(GP17)*

However, the interviews revealed that a diagnosis was not always possible or feasible resulting in diagnostic uncertainty. Diagnostic uncertainty might not necessarily translate into clinical uncertainty. For GPs, a referral was still clearly indicated by a positive test result or when symptoms met referral guidelines. Therefore, the diagnostic decision, in my data, became a question about how certain and confident a GP felt about managing the patient in primary care and when they needed to refer. GPs described their role in diagnosing patients as being the referral filter in the diagnostic process towards appropriate management.

"... it's about providing that filter so that those who have more IBD-like symptoms (...) are finding their way into secondary care for the appropriate management, whereas the IBS can be managed effectively in primary care...." (GP05)

5.5.1.1.2 Ambiguous clinical need

GPs discussed a number of situations that could result in clinical uncertainty of how to manage patients. These may arise when patients have uncertain symptoms,

when management fails and the working diagnosis comes under scrutiny or things change and GPs worry that they may miss something serious. In situations of clinical uncertainty, the clinical need for a referral may be less certain and in some cases may become ambiguous.

“And if we ourselves are also uncertain whether we have covered all the ground and whether we have got the diagnosis right? (...) Most of the things are evolving and there is a lot of uncertainty about and there is the potential to miss things and the implications thereof, both for the patient and ourselves as well. So I think when in doubt it is good to refer.” (GP08)

It became clear from my data that in some situations there were no hard criteria for a referral decision but it was the subtleties and the mannerism of the patient that concerned the GP sufficiently to prompt a referral. Only the outcome of the referral would be able to confirm the clinical need of referrals based on intuition.

“But there is something in there as well about the sixth sense, there is something in there about, ‘This just doesn’t feel right.’ And quite often, if you listen to your little voice it will get you to do the right thing. (...) There are hard criteria and then there are, ‘Mmm, what does this feel like? I think this needs to be referred.’” (GP11)

In this context of referral decisions under uncertainty, GPs discussed the appropriateness of referrals. GPs classed a referral, where the clinical need was confirmed, as a good referral.

“...a good referral would be where you were uncertain of the diagnosis and the consultant sort of like agrees with you and they feel that yes they need further investigation and then they have further tests etc.” (GP10)

On the other hand, they discussed inappropriate referrals in the context of missed opportunities to manage the patient in primary care, if the referral was purely motivated by the GPs uncertainty or in response to patient expectation, when the patient did not attend, or when the patient was discharged after one outpatient appointment. It became apparent from the data that it depended on the perspective

and context chosen whether a referral was classed as wasteful; that is, wasteful to the NHS, wasteful to the consultant or wasteful to the patient.

“So yes, it depends on what you consider a wasted referral is. And I suppose it is the perceptions of who is doing the referral of whether it is wasted or whether the consultant thinks it is a wasted referral. (...) I would be interested to see if a consultant would consider that [provide advice on management] as a waste or whether they see that as part of their role to give that advice.” (GP12)

5.5.1.1.3 Summarising clinical need for referral

My findings suggested that clear symptoms can promote diagnostic certainty which in turn may translate into a clear referral decision. Diagnosis, however, was not always possible nor important, for instance if GPs were confident in how to manage the patient. Therefore, diagnostic uncertainty may not necessarily translate into clinical uncertainty as appropriate patient management was the major concern of GPs and diagnosis was just a means to management. Under clinical uncertainty, however, patient management was unclear and a referral decision may be considered inappropriate. Referral was, therefore, either a consequence of the certainty in a serious diagnosis or the uncertainty in the management of the patient and depended on how confident the GP felt about managing the patient in primary care. When the clinical need for a referral was uncertain other influences such as the doctor-patient interaction and the doctor's personality and experience may carry more weight and become more influential on the referral decision.

5.5.1.2 Doctor-patient interaction

“Negotiation, it's what we do every single day, (...) we're working with the patient to re-educate them and it's a very complex procedure when you've only got a small amount of time, as a GP, you have to really pick up on patient cues, their concerns, their ideas, what their expectations are and try and work through from basic principles and I always feel that showing your working, explaining why you're saying certain things usually gets the patient on side.”(GP14)

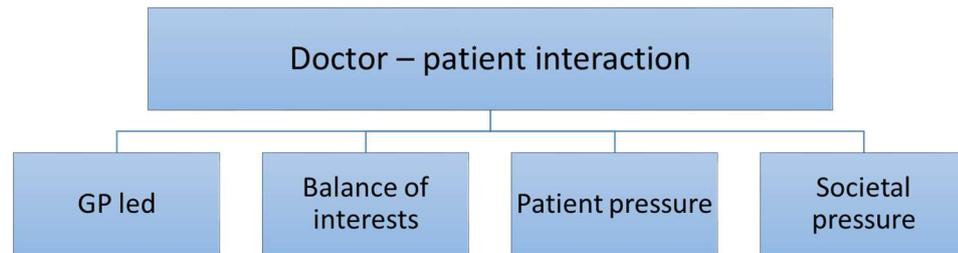


Figure 5.4 Theme 2 of referral influences - Doctor-patient interaction - and its subthemes

The decision to refer was influenced by how the GP and patient interacted with each other during the consultation based on their expectations, personalities, attitudes and beliefs. Because every GP and every patient differs in these aspects, GPs explained that every doctor-patient encounter is unique and it is the chemistry between the particular patient and the particular GP that can influence which direction the referral decision may take. My analysis suggested that the interaction during a referral decision between the patient and their GP can be dominated by the GP (**GP led**) or it can be dominated by the preferences of the patient (**patient pressure**). Furthermore, it can be more balanced based on patient choice and consent (**balance of interests**) or influenced by external factors such as a desire for tests and an intolerance of uncertainty in the population (**societal pressure**) (Figure 5.4).

5.5.1.2.1 GP led interaction

A GP led interaction was characterised by decisions made primarily by the GP in which the question dominated whether there was a clinical need for a referral. GPs described the importance of avoiding investigations and referrals if not indicated, using negotiation skills to explain decisions and being able to say no.

“But you know some patients come in and they can say things like “Oh just a quick one. I just need a referral”. And so you have to give them a fair go. But the starting point is yes well we’ll see about that. Because (...) I certainly don’t give patients everything they want. (...) And you just have to say “No you don’t need it”. (GP03)

Furthermore, GPs detailed how they put patients in charge of their symptoms, with an expectation that they learn to manage them themselves.

“Look, it isn’t anything. Let’s just leave it. I know it’s affecting you, but it’s not anything medically. Let’s try and tackle how you’re going to manage it on a daily basis rather than how I’m going to manage it on a daily basis because it’s not me that’s going through it.” (GP07)

5.5.1.2.2 Balance of interests

GPs also described a more balanced interaction with some patients, which was characterised by shared decision making, patient choice and patient consent. GPs reported themselves as aiming to empower patients by offering tools and support, respect the patients’ preferences and build a good relationship with the patient.

“I always say to patients my role is to put all the facts before you and then we come up with a solution together. (...) I think it is a conversation not a “this is what will happen.” It’s not a didactic thing, it is a conversation. You have to buy in the understanding from the patient as well.” (GP05)

My analysis suggested that responsiveness to the patients’ needs and underlying worries helps the GP to understand the motivation for help seeking and potentially unrealistic expectations. These may be met with a listening ear and counselling.

“I think the biggest challenge for a GP is trying to establish the patient’s expectations and what they expect from me when they come to see me. And trying to make sure that I have managed to listen to their concerns and what they are actually worried about. (GP15)

Providing patient centred care, as described by the GPs, relies on making management decision on a patient by patient basis. This also involved being professional and diplomatic in situations of disagreement. According to the GPs, this was key in order to enable them to find some middle ground and manage the patient long term.

“So if you really feel sure that the patient is asking for something that is not warranted but yet it is getting to a confrontation point then I try to be

diplomatic and work around that. Because you really want to try and keep the patient on board as far as possible as you are also aware that this is a condition that you are not just managing them for today but actually in the long term.” (GP06)

In terms of referral, this meant finding out what the patients expect from the referral and offering equivalent services in primary care.

“Most often, you work it through as a negotiation (...) “Tell me what it’s for. Let’s see if there’s anything we can do?” (...) “Okay, well it’s no problem to send you up to the specialist. If we do send you up, the first things they’re going to do is this. Why don’t we do that round of tests here?”” (GP09)

5.5.1.2.3 Patient pressure

In my data, referral decisions in response to patient pressure were a way out of a stressful situation to salvage the relationship with the patient. GPs perceived pressure indirectly from patients who did not actively hassle for a referral, but kept on wanting a solution for their symptoms and the GP ran out of options. One GP identified such referrals as similar to giving antibiotics for sore throats.

“Normally I try to stick to my guns but occasionally you have to give in. It’s like antibiotics for sore throats. That kind of scenario. Most of the time you can stick to your guns but sometimes you have to just for practical reasons of... It’s hard to explain why sometimes you do back down in such situations because just talking about it now it seems a completely irrational thing to do. You know if you are certain about something then why would you back down but sometimes I suppose it’s the short time of the consultation and eh the relationship you have with patients. So there are lots of different factors involved.” (GP01)

GPs explained that these referrals aimed to relieve the patient’s suffering as an empathetic reaction to the patient’s anxiety and state of mind.

“I think it’s just if somebody is particularly if you feel they are particularly... I mean in this specific scenario if they got specific worries quite often people mention cancer. That’s something they are worried about. And if people are clearly you know upset and very stressed about it then I’d be more likely to

do some more tests for them even if I was quite sure that it was IBS. (...) It's more an empathetic kind of thing." (GP01)

They also admitted that the referrals provided relief for the GP.

"The positive (...) is that they have stopped coming to see me for a while which is kind of respite for you because you think you will get a breathing space from them coming down every other week. So it gives you a bit of respite and that is an awful thing to say because there aren't many you can't manage but there are one or two where you do feel you need a little bit of a break from trying to come up with ideas and seeming particularly interested in their IBS." (GP18)

My analysis suggested that patients could also openly express a preference to see a specialist when the GP cannot offer a response sufficiently rapid or offer objective evidence that would be needed to satisfy the patient. If the patient expectations were not met and the GP failed to reassure the patient, the patient might actively pressure GPs into a referral. Some patients may presume that it is their right to be referred and use the GP as an instrument into secondary care.

"So some patients want a referral. They want to see an expert. They come in and that's their goal of the consultation." (GP03)

5.5.1.2.4 Societal pressure

According to my data, interactions between GPs and patients could also be influenced by a general awareness of societal views on patient rights and legal actions, which could translate into pressure on GPs decision making. GPs witnessed the effect of an expectation for investigations in the general population and an intolerance of uncertainty and mistakes resulting in medico-legal pressure to practice more defensive medicine. Within this climate it was thought that GPs feel it might be safer to refer than take the risk.

"I think the landscape of medicine has changed in the time I have been doing this job in that going back some twenty years patients were much more tolerant to the doctor doing his best and the doctor forming an opinion based on the information he had. And although you would be right most of

the time there would be a small number of times where you would not be right. And patients were much more likely to accept that as the uncertainty of life. Now patients are much more 'where there is blame there is a claim' and are much less tolerant of uncertainty." (GP11)

5.5.1.2.5 Summarising the doctor-patient interaction in referral decisions

On the basis of what GPs reported, two seemingly identical clinical problems in two individual patient encounters could result in substantially different decisions for the management of the patient. This was due to the different types of doctor-patient interactions which determined the course and outcome of a consultation.

Furthermore, a referral on the basis of a patient's preference could be described as giving in to patient pressure, either real or perceived, or as a patient's right at the heart of patient centeredness and patient choice, reflecting two different types of doctor-patient interactions at play. Which interaction dominated appeared to depend on how the GP considered the patient and situation in which the request was made. Agreement between the GP and patient was more likely to facilitate shared decision making, while a disagreement required negotiation. The type of interaction that would operate during a patient consultation would clearly depend on both, the GP's and patient's characteristics as well as on the severity of the patient's symptoms. There were, therefore, strong links between the theme doctor-patient interaction and the themes of GP factors and clinical need. Exploring the patient characteristics independently would have added useful and interesting data but would have required additional interviews with patients.

5.5.1.3 GP factors

"I think some of it will be perhaps down to the personality of the GP and how you are able to handle conflict. Some GPs may feel better able to manage a conflict with patients than others and have a higher or lower threshold for actually having a dissatisfactory consultation." (GP06)



Figure 5.5 Theme 3 of referral influences - GP factors - and its subthemes

In my data, several GP specific factors influenced the outcome of a referral decision, which reflected the GPs' individuality. These included the **GP's mind set**, their **experiences** and their **motivations** (Figure 5.5). They determined how a GP might respond to the challenges and situations presented by different clinical presentations, different patients and the healthcare system. As a result of the differences in GPs' mind sets, experiences and motivations, I would expect a wide variation in referral decisions.

5.5.1.3.1 GP's mind set

While GPs are trained medical decision makers, my data suggested, that their decisions might also depend on the kind of person they are. Their personality may determine how GPs handle conflict, whether they may feel undervalued by a referral request or diminished in their professional pride when they have to make a referral.

“And I almost, sometimes I feel a slight sense of disappointment that I've got to refer on because I can't manage it or can't sort it out myself, that's a sort of professional pride thing.” (GP16)

Some GPs discussed personal challenges as an influence on referral decisions such as remaining objective when dealing with a returning patient or managing more difficult patients. GPs also admitted that their frame of mind on the day can impact a referral decision, reflecting that some decisions might be made by the GP as a person and not purely as the infallible clinician as is often expected.

“I recognise that some of them [IBS patients] can be very, very complex. Most aren't. Most are fairly straightforward but there are a complex few in there that can be you know, difficult to manage. And that is when I would use a gastroenterologist for support.” (GP18)

“And especially at the end of the day when you've had... when I am a bit fed up and I'm less sympathetic you know the IBS patient is your 'uh', this is not going to be easy. So it's those kind of uncharitable thoughts.” (GP03)

5.5.1.3.2 GP's experience

The second aspect of the GPs individuality, which emerged from the data, was the GPs' experience. My data suggested that GPs can learn from referrals and may

become more efficient in their decisions and more confident about their reasons for referral. With growing experience GPs apparently combined their experiences with common sense and insight which they used to evaluate past experiences, generalise encounters and form new referral decisions.

“Only fools are certain’ (...). And also the more you learn the more you don’t know. (...) So I am much more likely to refer but I guess what I have probably evolved into is that I know why I am referring more rather than just referring more because I do not know what it is.”(GP11)

GPs recognised that they have different referral thresholds and that GPs who had legal action taken against them, are more likely to be higher referrers for some time after.

“Once you’ve had legal action taken against you (...). Suddenly, you become more defensive in your medicine. So if I’d missed something like a... I thought it was an IBS not cancer I’m sure I would be much more careful and I’d probably refer more.” (GP03)

My analysis also suggested that GPs grow with experience. GPs talked positively about referral when their referral expectations were met. These were seen as establishing a diagnosis, input into management, and support for their management decision. GPs expected referrals to be more involved for IBD than IBS patients. For themselves, GPs hoped to get support from a referral in terms of reassurance, back up, confidence, respite, consolidation and restoration of their authority to manage the patient in primary care.

“One consultation of that nature might be enough to then allay their concerns, or almost give you as a GP then a little bit more authority to manage that patient and them to have the confidence.” (GP05)

They also referred to share responsibility for patients they worry about.

“I think we dissolve a bit of responsibility as we refer, because some of these patients have got things and we’re like, I just don’t know what to do. Actually, that can be quite stressful (...). But I think sometimes the patients I’m worried about, especially, referring them on makes me kind of go, okay, so

actually I don't know what's wrong, or I can't deal with this so I'll get someone else to see it, (...), someone else is going to take control.” (GP04)

However, GPs also encountered negative consequences of referral and negative experiences with referrals. They recognised that referrals aggravate patients' worries and create uncertainty while patients wait. GPs identified drawbacks of referral to be the subspecialisation of clinicians who were seen as not treating the patient as a whole person and testing automatically rather than as indicated. GPs associated negative consequences such as over investigation, iatrogenic complications, incidental findings and unnecessary treatment with some referrals. Some worried that patients would not feel the need to see their GP when under the care of a consultant. Furthermore, GPs talked about their experiences with dismissive consultants and the experience of referrals not living up to expectations.

“The downside is sometimes it doesn't quite live up to expectations. They are seen in a very busy clinic by a junior registrar who [is] actually just wanting the more serious patients.” (GP02)

In my data, GPs weighed the positive experiences against the drawbacks of referring into secondary care in their decision whether to refer a patient and when to refer a patient.

“... it takes judgement as to quite when you refer. You don't want to get right to the end of your line [of treatment options] as you are also aware that you have then six, eight weeks maybe even, before they see a specialist.” (GP05)

Some GPs described how they experience a referral as letting go of patients and secondary care taking over and described the care patients receive as sequential rather than shared.

“The consequences of the IBD referral is a little bit of letting go. You know, you have worked this patient up, you have made a presumptive diagnosis and hopefully reasonably quickly. You have explained it and then sometimes once they end up in the clinic you don't see them again for ages. (...) And whilst they get monitored up there you do sometimes feel that they have kind of taken over. So it's a bit of 'hang on a minute'. I am responsible for their

health and you are just responsible for their bowels. (...) So that is a negative thing about that relationship you have with the patient.” (GP18)

5.5.1.3.3 GP's motivation

The third aspect of the GPs individuality to emerge from my data encompassed everything that motivates GPs; their ambitions, their level of satisfaction and their perceived role. In the context of referral decisions, GPs discussed their ambition of not wanting to miss an opportunity to refer, but also to keep IBS patients out of secondary care, who can be managed effectively in primary care.

“If we could do some of the tests here and make those decisions and know where to go with them, that would be a really useful test to do here. I think just anything to keep people out of secondary care is a good thing, and we've got a good relationship with the patients, so we can test them and we can tell them the results.” (GP04)

GPs described satisfaction with the care they can offer as motivating. They listed a good level of patient knowledge, helpful feedback from consultants, availability of diagnostics and creative thinking to distinguish their clinical decision making from simple flow charts as enablers for satisfactory care.

“We have guidelines for so many things that it is nice to have a little bit of leeway to do referrals that you just have a feeling about.” (GP03)

Yet they communicated frustration with the referral system and the long waiting times leaving them feeling unsatisfied with the care they can give to their patients.

“...there's nothing available on Choose and Book because it just says to refer to the provider and you just refer them into this void, which you know can take nine months and (...) so some of the patients elect to go privately even though they may not be able to afford it and I always regard that as unsatisfactory. I certainly back the patient up in that and that's the choice that they want to make, then I'll arrange that but it is obviously, I feel downhearted that the NHS has let these people down so they have to fork out and go privately.” (GP16)

GPs described their role as managing the patient. They believed this would not always necessitate a diagnosis in primary care but passing patients along the diagnostic pathway. GPs stressed that they are generalists, and refer on to specialists after good history taking and examination.

“I’m not an expert in Gastroenterology, (...) but I hope I’m expert enough to address that and then forward people onto the right parties. So, I’m more than happy for them to go to secondary care for them to get the proper, full job.” (GP17)

The view on the GPs’ gatekeeping role was divided amongst interviewees. While some GPs identified themselves as gatekeepers for secondary care, others did not like to be seen as the triage person for secondary care.

“Yes they always say that GPs are gatekeepers for secondary care. I think...I don’t know how to answer that question really. I do not like to be a triage person for the secondary care, I’m not a triage person for the secondary care, I’m a clinician with my independent practice really. But I think secondary care is there for me to support me with complicated cases. I’m not just sitting here and saying that can go to gastro that can go to surgery and that... I’m here to make my own judgement and secondary care is to complement my practice really.” (GP19)

5.5.1.3.4 Summarising GP factors in referral decisions

My interviews revealed that GP factors, which were thought to influence referral decisions, were very personal. GPs described individual personalities, challenges, experiences and motivations. Their personalities shaped the way they grew with experiences, both positive and negative, and how they experienced referrals. Their mind set on the day was reported to influence how they manage patients. Their perceived role determined how they viewed the referral system. Their experiences with legal action, consequences of referral, individual consultants and the organisation of care affected what GPs expected from referrals for themselves as well as for the patient. Considering the different personalities and individual experiences, referral habits of GPs could vary greatly.

5.5.1.4 Healthcare system

“If the service was there like it is in some countries, you know, so if we can get a Gastroenterology opinion about you, this week, we would probably do that more often, wouldn’t we, than if it was a nine month wait and so we would, by definition, almost certainly refer more patients and that’s really interesting because that’s just the service provision influencing our decision making, but it would be a much more expensive system...” (GP16)

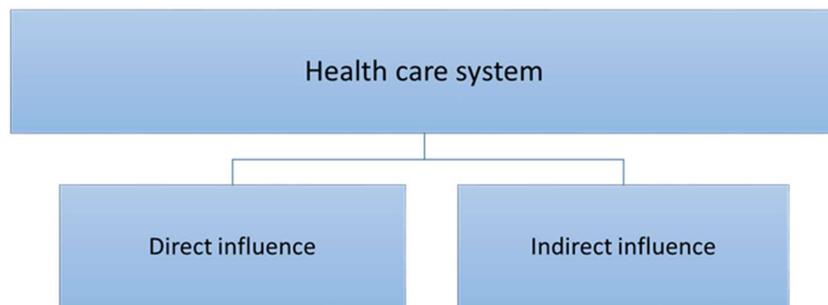


Figure 5.6 Theme 4 of referral influences - Health care system - and its subthemes

My analysis suggested that the healthcare system and how it is financed, organised and delivered to the population can influence GPs’ decisions directly and indirectly (Figure 5.6). **Direct influences** consisted of the organisation of care which encompasses guidelines and pathways, the national referral system and services which provide advice and guidance and the pressure exerted by CCGs on GPs referral rates. **Indirect influences** included the relationship with secondary care and the availability of resources including the long waiting times for secondary care appointments in the NHS.

5.5.1.4.1 Direct influence

One direct influence of the health care system on referral decisions in my data was the organisation of care. This included services and decision aids available to GPs and how GPs accessed them. GPs described guidelines and pathways as a major influence on referral decisions. GPs felt that guidelines make it safer for them but also take away the challenge they enjoy. Without guidelines GPs would very likely show more variation in their referral decisions. In the following quote GP03 describes ambivalence about guidance for referral:

“They either meet the criteria or they don’t. And if they meet the criteria we just refer them. Whereas these there is not necessarily criteria for when to refer to gastro. So I imagine different GPs probably my partners and me and at different times we have slightly different... It’s more subjective isn’t it? (...) Because we have guidelines for so many things that it is nice to have a little bit of leeway to do referrals that you just have a feeling about. Or is that about...is that...would it be better to have guidelines for everything and clear cut and you can defend yourself by saying I referred according to the guidelines.” (GP03)

GPs described several services that provided support for their referral decisions. Formalised services by gastroenterology departments included an e-mail advice service called advice and guidance as well as dedicated phone lines provided by some hospitals. In addition to specialist advice, GPs explained using peer support to discuss uncertain cases. Finally, GPs who have personal relationships with consultants mentioned phoning or writing informally for advice.

“I do increasingly write to consultants just to ask for advice on a specific matter with regards to a patient where in the past I may have referred them.” (GP01)

GPs were aware of the expectation to refer all patients through the e-Referral System (previously Choose and Book), however, the shortage of timely appointments in gastroenterology forced them to revert back to local ‘letter’ referrals, and some opposed being forced to use it. They also criticised the system for disjointed services, lack of available appointments and loss of the personal relationship with local gastroenterologists. In the context of access and waiting times, GPs discussed making three types of referrals: Two-week cancer referrals, urgent referrals and routine referrals.

“The two-week pathway is a really rigid pathway but other than that it is a letter which is dictated and sent. And there is the Choose and Book system which secretaries use to try and insert patients into available appointments but often there will be no appointments free for months and months and months. So if you can dictate it and send it urgently or fax it etc., then that consultant in secondary care has the ability to perhaps slot a patient into a clinic.” (GP15)

The second direct influence of the healthcare system on referral decisions was described by the GPs as the pressure the healthcare system placed on them to avoid referrals. This can put referral rates at the forefront of GPs' minds with the result that GPs felt the need to be able to justify referrals to themselves, the patient, colleagues and consultants. The pressure came first and foremost from the CCGs, who publish referral charts to name and shame high referring practices.

"I'll say one consequence [of referral] is another goes against my name in the column for referring. And we are... Especially this year we are being really put under the thumb to not refer as much. (...) we are measured on how much we are referring and compared with the other practices locally and nationally and fingers pointed if we are referring too much. ...because we are under pressure not to refer as much we all think about referrals more." (GP03)

On the other hand, GPs voiced concerns over not referring enough and missing or delaying cancer diagnoses and consequently being scrutinised for low pick up rates.

"Luckily we are a relatively low referrers but you don't want to be too low because if you are right at the bottom that means you are not referring enough and you are missing things." (GP03)

While for some GPs the referral rates and their referral reputation was constantly on their minds others were not concerned by referral rates, which they felt they can justify. They disregarded the pressure from the CCG and claimed to refer on symptoms alone. This illustrated how the GPs' personality shaped the influence of the health care system on referral decisions of individual GPs.

"I mean I hope the CCG doesn't influence my referral decision making at all, or to any great degree because if a patient needs to be referred then I hope I would refer that patient on. (...) That is more likely to affect your referral pattern than the CCG to be honest with you." (GP13)

5.5.1.4.2 Indirect influence

My findings suggested that the way in which the health care system delivers its services to the population has an indirect impact on the decision to refer. GPs praised the link with IBD specialist nurses who absorb some of the delays of

referrals to gastroenterology and specific IBS clinics which can avert referral to gastroenterology. Furthermore, GPs felt that the consultation time can influence their decision making.

“And I think there is the challenge within general practice to give patients time to tell that story. And I think that is a particular problem at the moment. If you give the patient long enough they will tell you the answer themselves. (...) but we are losing that ability to give that patient time.” (GP11)

However, availability of resources varied depending on which hospital the practice was linked up with and which CCG area they were in. Availability of resources appeared to lead to variation in referral habits. GPs reported that they refer fewer because specialist access is poor, offer patients to go private and would use gastroenterology services more often if appointments were timelier.

“And I try not to refer them. (...) because I just think, ‘They are going to wait forever and are they really going to be getting any better opinion than they would be getting from me? I think not.’” (GP18)

The inconsistent availability of resources also influenced the GPs’ relationship with secondary care. GPs described how they work the system to try and achieve the best care for their patients. GPs resorted to using the advice and guidance service to try and squeeze an urgent patient in, used their personal relationships with consultants to accelerate an urgent referral and to get around the referral support system, they asked secretaries to chase up a referral and exploited the two-week cancer pathway to achieve a prompt investigation.

“It is easier to get to see the surgeons so it is easy to get a test. I can get an endoscopy done quickly and I can get a colonoscopy done quickly. So I think people bypass the system by manipulating the access to the tests. So for example if you have the older age group it is very easy to get a two-week wait: change in bowel, tick the form, done.” (GP18)

GPs expressed mixed views about their satisfaction with gastroenterology services and consultants. They felt that the relationship with consultants was dwindling. They believed a good relationship can facilitate urgent referrals, the writing of high quality

referral letters and access to advice. They also described frustration about being left to deal with some irrational decisions made in secondary care.

“The hospital play games at times so sometimes we ask for a direct access colonoscopy and they see the consultant in clinic before they have the slot and sometimes they do it. (...) Not always what you ask for is done (...). Sometimes they’ll get more money if they’re seen as an outpatient appointment than having an investigation. (...) Sometimes it’s our choice and sometimes it’s not our choice with how the referral pathway works. It’s quite complicated unfortunately. I wish it could be simpler, but it’s not sometimes.” (GP07)

5.5.1.4.3 Summarising the influence of the health care system on referral decisions

The influence of the healthcare system on referral decisions, according to my analysis, was twofold. Firstly, the healthcare system influenced referrals through active measures which were put in place to control the flow of patients. These included guidelines, pathways and referral charts. Secondly, the health care system has an influence through the way its services are delivered, that is, what is provided and at what time scale. This had an impact on the GPs views they reported on the referral system and secondary care services. However, there is variation in how GPs respond to the pressure and influences from the health care system. The level of impact on the referral decision, I believe, is, therefore, variable.

5.5.1.5 Summarising the influences on referral decisions

My interview analysis showed that referral decisions are complex. They can be influenced by a number of factors which I described as clinical need, the doctor-patient interaction, GP factors and the health care system. According to my data, referral is either a consequence of the certainty in a serious diagnosis or the uncertainty in the management of the patient. In cases of clinical uncertainty, the doctor-patient interaction and the doctor’s personality are likely to become more influential on the referral decision than clinical need. The GP’s personality (as well as the patient’s characteristics) appeared to determine how GPs interact with their patients. This can determine whether decisions are dominated by the GP, by patient preferences or by shared decision making. In addition to the GP’s personality, decisions may also be influenced by GP’s individual challenges, experiences and motivations. GP factors also seem to affect how GPs respond to pressure and

influences from the health care system including guidelines, referral charts, waiting times and secondary care services. According to the data, the influences are interrelated and it is the product of all influences that determines the outcome of individual referral decisions. According to the GPs, patient assessment is a process of gathering pieces of information in an environment that is greatly influenced by the GPs' characteristics, knowledge and experience as well as their interaction with the patient and the health care system.

5.5.2 FC TESTING WITHIN THE REFERRAL DECISION

Tests like the FC test are believed to support GPs' diagnoses and impact their referral decision with a positive effect on referral rates.⁵⁹ However, this view assumes that GPs use tests only as diagnostic aids. In the following sections, I report my findings on the role GPs assign FC testing in their clinical practice and on how the FC test influences referral decisions. I start with a description of the outcomes of FC testing, both positive and negative, and how they interact with the referral influences (section 5.5.2.1). The descriptive arrows in Figure 5.7 illustrate how the outcomes control the four main referral influences. In addition, I provide an overall evaluation on whether the referral decision process is likely to be supported or unsupported by FC testing or whether FC testing might have no impact on referral decisions at all. Considering the different directions of effect on referral decisions, the overall impact of FC testing on the referral rate is difficult to predict and will depend on who GPs consider for testing. Therefore, we also need to understand the GPs' perception of the role (section 5.5.2.2) and place (section 5.5.2.3) of FC testing in primary care and GPs' requirements for testing (section 5.5.2.4). Taken together, this helps us to understand the role of FC testing in primary care.

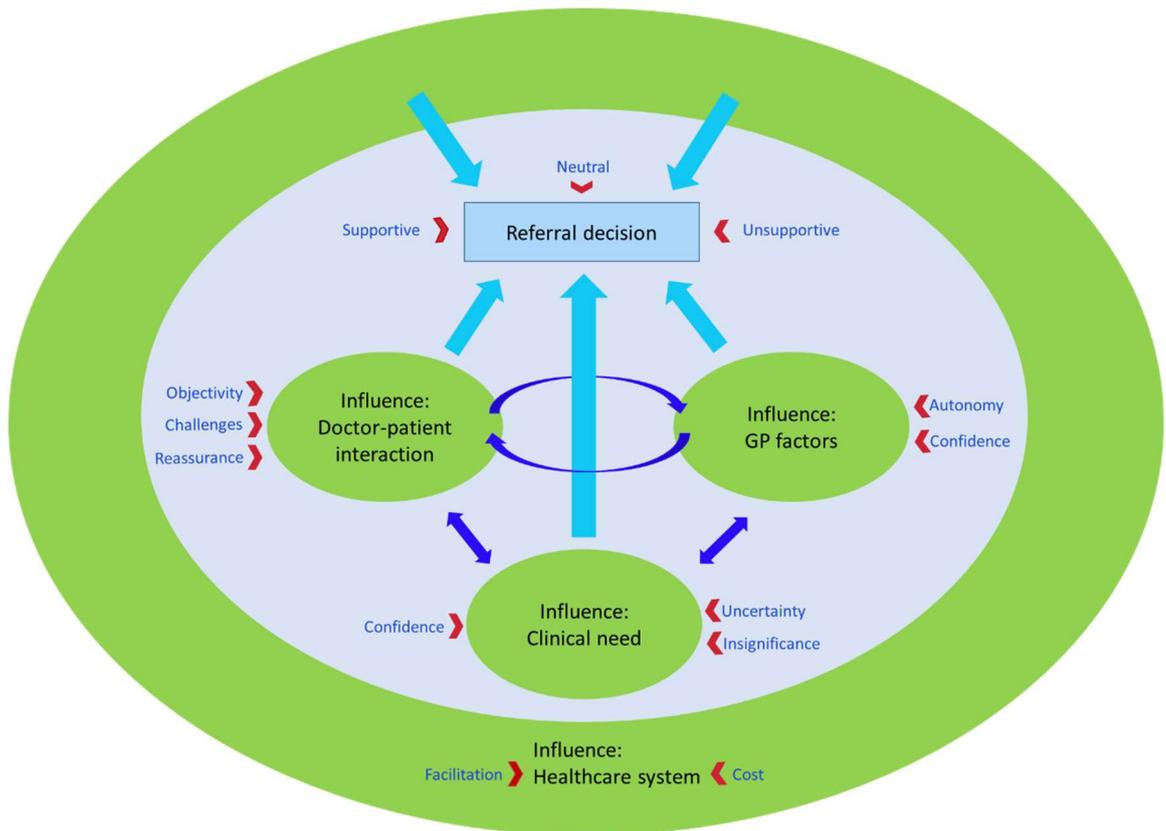


Figure 5.7 Overview of referral influences highlighting the outcomes of FC testing and the direction of their impact on the referral decision

5.5.2.1 Outcomes of FC testing

“Then I reassure the patient that we have a negative test but that doesn’t mean that we don’t..., with anything we’ll safety net, and nothing is absolutely concrete. (...) one test that’s showing it’s unlikely for something but you can never say never, with anything.” (GP17)

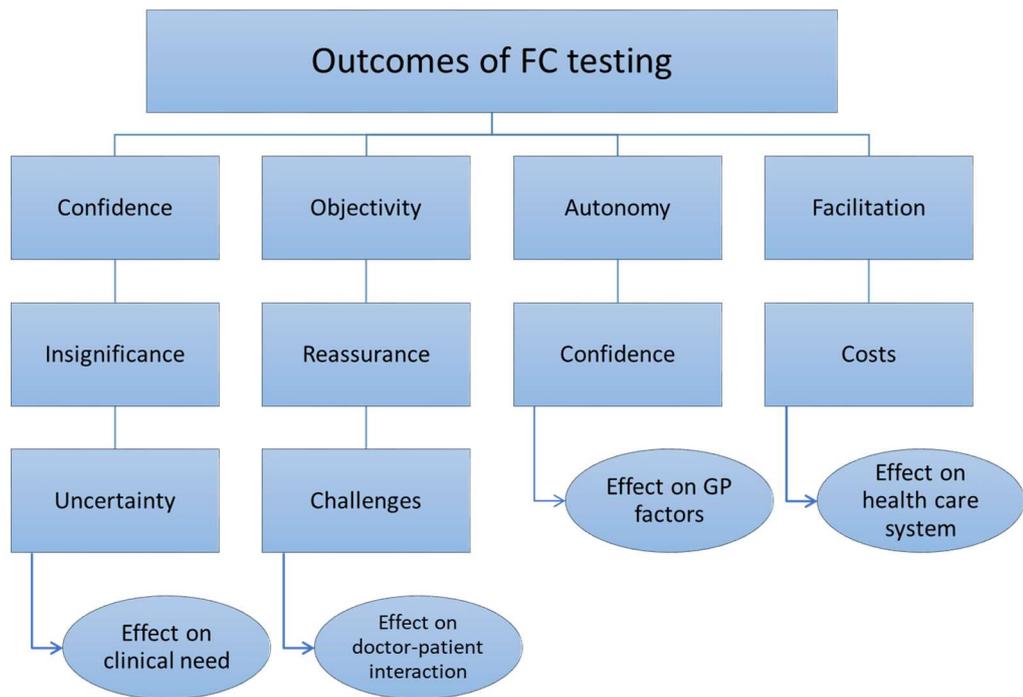


Figure 5.8 Overview of outcomes of FC testing and their effect on the four influences of referral decisions

5.5.2.1.1 Outcomes of FC testing affecting clinical need for referral

FC testing can affect perceived clinical need for a referral in various ways (Figure 5.8). From my data, it appears that FC testing can confer **confidence** in a diagnosis in primary care and subsequent management decisions, it may be **insignificant** in changing prior belief and it can convey **uncertainty** when test results are difficult to interpret or results are contradictory to prior belief. As a result, information gained from ordering an FC test can be supportive, neutral or unsupportive of the referral decision process depending on which situation is at hand.

FC testing was reported to confer confidence in a working diagnosis and subsequent management decisions. The level of confidence depended on the level of agreement between the GP's working diagnosis and the test result. A negative test result reinforced a working diagnosis of IBS and GPs felt more confident not to refer. A positive test result in combination with concerns over IBD prompted an earlier referral. A positive result changed the equation irrespective of the GPs prior perception of the symptoms and almost automatically generated a referral irrespective of the possibility of a false positive.

“And obviously, similarly if it’s positive then that’s a worry. It’s not absolutely diagnostic. It doesn’t mean that they definitely have IBD but it is certainly one of the concerns now. So you would certainly refer if it was positive. So it is a very clear kind of tool for helping in that decision.” (GP01)

GPs stressed that while a negative FC test was helpful for decision making it did not automatically rule out a referral because of the worry of missing a serious condition if symptoms persisted. If concerns over symptoms were sufficiently high GPs referred irrespective of the FC result. When the clinical need for a referral had already been established, the usefulness of testing in such situations was questioned as the test was insignificant in changing prior belief.

“So a patient who has got very good symptoms for inflammatory bowel disease I am probably going to refer even if the test is negative, which begs the question why do the test I suppose.” (GP11)

FC testing caused uncertainty when results were ambiguous. GPs were aware that the FC test does not provide a definitive answer, cannot exclude all conditions, may provide false positives and false negatives, may be less reliable in the older age group and can produce indeterminate results. While the feeling was that a slightly raised test was still abnormal, a definitive management decision could only be made following a repeat test.

“I think it is the borderline tests which I would imagine which would raise more uncertainty.” (GP12)

5.5.2.1.2 Outcomes of FC testing affecting the doctor-patient interaction

FC testing can assist or impede referral negotiations between patients and GPs. My analysis suggested that FC testing can provide **objectivity** and **reassurance** in managing patient expectations and may be supportive in the referral decision process if reassurance is in line with the working diagnosis. FC testing can also be unsupportive of the referral decision when it creates additional **challenges**, for instance, when patients overestimate the test’s diagnostic ability.

FC testing was described by GPs to have a positive effect on the relationship with patients. GPs used the FC test as a tool to communicate diagnoses and

management decisions and felt that it provided more objective evidence in the management of difficult patients.

“The last one I remember, which sticks in my mind, was the young girl who came in and out with what everyone thought was IBS, they thought she was, they actually thought she was faking her symptoms because she had depression anxiety, then one of my colleagues actually did the Calprotectin and it came back as 430, I think it was, it was pretty high and it turns out that she did have IBD.” (GP14)

An FC test can be reassuring if the results are in line with the working diagnosis. It can reassure the GP that they have not missed any serious disease and GPs can pass on this reassurance to patients. With this in mind a negative FC test was described as the most useful but none of the GPs indicated that using FC testing for reassurance changed their referral decision.

“You’re thinking, ‘This is almost certainly IBS but some of the features are not, are a little bit more concerning and, you know, I’d hate to miss Crohn’s Disease or ulcerative colitis’, and that’s when the test is really useful. You can send it off and it’s thoroughly normal, (...) then that’s really reassuring and you can tell the patient, you know, ‘It’s normal’.” (GP16)

FC testing brought about additional challenges when patients saw test results as black and white. One GP explained that patients placed more importance on a test result than GPs, which can cause disagreement. Testing was thought to inevitably create additional worries in patients about the test itself and the seriousness of their symptoms. Patient expectation and patient worry may result in referral where it wasn’t initially considered.

“Because we all like tests results, they look good on paper. Where there is a piece of paper with a number written down on it, it is very clear cut and of course it is not.” (GP11)

5.5.2.1.3 Outcomes of FC testing affecting GP factors

FC testing may provide a tool to GPs to meet their ambition to keep patients out of secondary care by giving them **autonomy** in testing and managing patients in primary care. According to the GPs, FC testing in primary care gave them more

confidence in interpreting test results and ownership of the management of patients. GPs felt enabled to put some temporary measures in place based on test results for patients waiting for their secondary care appointment.

“So I guess it [FC testing] gives you flexibility in primary care and it helps you to make management plans a bit more proactively really.” (GP19)

5.5.2.1.4 Outcomes of FC testing affecting the health care system

FC testing may **facilitate** timeliness of management and referrals and may relieve the referral burden. FC testing may be supportive in the referral decision by relieving some pressure and tension GPs described working within the current health care system. GPs believed that FC testing had the potential to reduce delays because testing in primary care happened sooner, it added weight to a referral making it more urgent and improved the quality of referral letters.

“It [FC testing] might mean the referral has a bit more weight to it and gives a clearer indication that the patient perhaps needing to see gastroenterology. So it may be a positive thing in terms of properly funnelling patients correctly into the right specialty and being seen in a timely manner by the consultant.” (GP15)

FC testing may also lead to savings to the health care system by reducing the number of invasive investigations. GPs were critical, however, that unnecessary testing could be **costly** to the health care system.

“I think testing has got its place in primary care, it’s just using it for the right patient at the right time and targeting it so that you’re not just doing a suite of investigations. (...) last year (...) we found that we were spending £1 million a year on LFTs [Liver Function Tests] and about another half a million on CRPs [C-Reactive Protein]. Actually, it was just because we were all so used to ticking the box...” (GP07)

5.5.2.1.5 Summarising outcomes of FC testing

My findings suggest that outcomes of FC testing, as described by the GPs, may have limited direct impact on the referral decision. This is aside from a clear positive FC test, which nearly always resulted in a referral. GPs reported that with FC testing

they were more confident not to refer in some cases. However, a negative FC test did not automatically rule out a referral. GPs used FC testing as a tool to communicate diagnoses and management decisions and to reassure patients. However, none of the GPs indicated that using FC testing for reassurance changed their referral decision. In my data, FC testing affected all influences on the referral decision and shaped the referral decision indirectly. As a supportive tool, it has the potential to decrease referral rates by providing confidence and timeliness of patient management. On the other hand, it can be unsupportive and result in disagreement and uncertainty creating referrals. Finally, FC testing can be initiated for reassurance in patients, for whom referral was not considered to begin with.

5.5.2.2 Role of FC testing

“Initially it’s, it’s tricky to say whether or not I’m trying to rule in or rule out IBS, because I know that it’s not, it’s mainly a tool to pick up inflammatory disease as opposed to rule in IBS but it is, it is a useful tool, particularly in these groups of patients, to show them that we are sort of looking at a lot of different things.” (GP17)

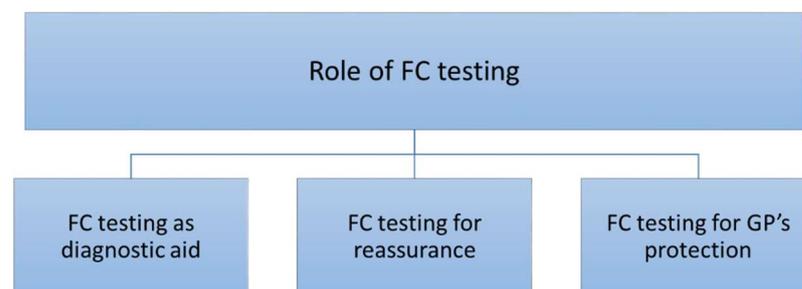


Figure 5.9 GPs’ interpretation of the role of FC testing in general practice

My analysis suggested that the role of FC testing is defined by the underlying reasons why GPs use FC testing (Figure 5.9). GPs used FC testing for three main reasons. First, they looked to FC testing as a **diagnostic aid** in cases of diagnostic uncertainty. Second, they used FC testing to **reassure** patients or find reassurance for themselves. And finally, GPs considered that FC testing may provide **protection** from medico-legal repercussions from disputed referral decisions.

5.5.2.2.1 FC testing as diagnostic aid

GPs valued the additional evidence the FC test can provide to either detect or exclude IBD. GPs felt that it is a useful tool to address diagnostic uncertainty in

cases where symptoms were not well defined or when a blood test came back slightly raised. In these cases, GPs assigned FC testing a role in ruling out IBD.

“I think if a patient perhaps had some symptoms that sounded a bit like IBS but they had been perhaps a little bit more severe, the diarrhoea was a bit more frequent or there was a bit of a history of rectal bleeding intermittently so in a bit more of a grey line it [the FC test] could be useful then.” (GP15)

By ruling out IBD, GPs also used FC testing to more positively diagnose IBS. With FC testing, GPs felt more confident in giving symptoms time to develop or in delaying a referral.

“If it [FC test] comes back negative on one where I was feeling fairly confident anyhow, it would strengthen your confidence in not escalating, as long as you’re still listening out for and keeping the door open.” (GP09)

GPs also found the test useful in confirming suspected IBD and make an earlier referral. Some ordered the test before patients went for colonoscopy to add weight to a referral.

“The faecal calprotectin, (...) using that as a filter or an extra back-up to try and facilitate a timely referral. If I feel that you’ve got the blood test, the inflammatory markers are up and if you had the faecal calprotectin you’ve got all that there, hopefully then when you make the referral they’ll be seen as soon as possible.” (GP05)

Finally, GPs used the test to differentiate between IBS and IBD where symptoms are suggestive of either IBD or IBS.

“These [FC tests] are for the patients who have reasonably prolonged symptoms or symptoms significantly suggestive of either IBS or IBD and trying to sort out that group, isn’t it?” (GP16)

While these situations describe effectively the same decision, that is, whether to manage or refer, GPs had slightly different patient stories in mind and phrased expectations accordingly. It is these nuances that reveal the individuality of each decision and how tests as decision aids may need to adapt.

5.5.2.2.2 FC testing for reassurance

GPs ascribed FC testing a role in reassuring patients as well as GPs by confirming the absence of inflammation and placing confidence in the decision to manage the patient in general practice. While FC testing may be used to reassure patients, GPs appreciated that they cannot investigate everybody and that it should not be used just for the GPs' contentment. One GP expressed ambivalence about FC testing for reassurance.

“Reassure patients where I think it's IBS. I think so, mainly that's why we use it [FC test] here. (...), it should be really using it for what you suspect is inflammatory bowel disease. But unfortunately sometimes it's just reassurance that you use it for.” (GP19)

5.5.2.2.3 FC testing for GP's protection

In my interviews, it became apparent that some GPs felt unsupported by the NHS when things go wrong. These GPs assigned FC testing a role in providing additional evidence to back up a referral decision. One GP commented that completing an FC test can protect the GP because they can prove that they have done everything recommended in the guidance in case a referral decision may end up in court.

“There is a group of people that you are pretty sure that you know what it is. You are almost certain you know what it is. And if I am honest with you, you probably might not refer anyway. If you are giving me a test that I can do which confirms the fact that they do not need referral and it is a little bit more evidence to protect me and the patient then yes, I will do that test. (...) NICE is an organisation which offers guidelines. (...) if I am standing in court and I am trying to defend myself for not referring on an inflammatory bowel disease that guideline will be used against me.” (GP11)

5.5.2.2.4 Summarising the role of FC testing

My data showed that GPs look to FC testing mainly for diagnostic certainty and reassurance to support the decision whether to manage the patient in primary care or to refer on. FC testing also provided confidence that nothing had been missed and that guidance had been followed. There were nuanced differences in the

expectations GPs had of the test in terms of its role as diagnostic aid and the appropriateness of using testing for reassurance.

5.5.2.3 Place of testing

“If they were coming in with more symptoms of irritable bowel you might be thinking of just doing some simple blood tests initially rather than, you wouldn’t necessarily leap in with a faecal calprotectin but that might make up part of your investigation. (...) If they have definitely got symptoms that are suggesting inflammatory bowel you would probably at that stage do that test along with all the blood tests and be thinking about early referral.” (GP18)

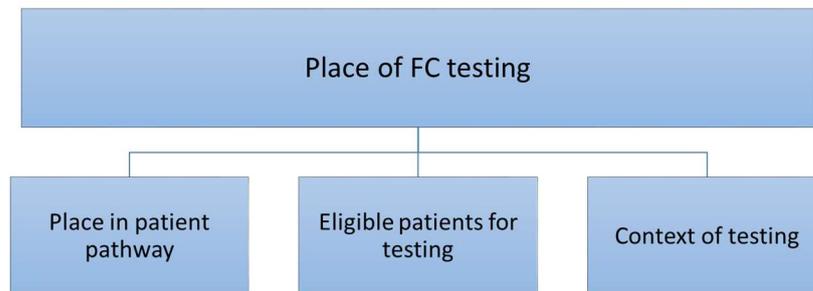


Figure 5.10 GPs’ views on the place of FC testing in general practice

The place of FC testing (Figure 5.10) was closely linked to the role GPs ascribed to FC testing. The reasons for testing determined when GPs considered the test to be useful in the assessment of patients. Using the GPs accounts the place of FC testing can be described in three ways: first, its place in the **patient pathway**; second, the definition of **suitable patients** for testing; and third, the **context** in which GPs test.

5.5.2.3.1 Place in the patient pathway

Some GPs described FC testing as a second line test following blood tests in the patient pathway of IBS/IBD symptoms. For other GPs FC testing was becoming more of a baseline test alongside blood tests. Whether FC testing was classed as a second line or baseline test followed from the severity of symptoms in patients who GPs would consider appropriate for testing.

“We tend to investigate all patients at least with a full blood count and with a coeliac antibody screen. I think the threshold for screening for faecal

calprotectin is becoming more regular now as well. So they would be the baseline tests I would tend to do.” (GP06)

5.5.2.3.2 Eligible patients

Eligibility for testing was often described in terms of patient demographics or symptoms meeting local criteria including age and duration of symptoms. GPs tested patients with new symptoms presenting for the first time but also patients with persistent and worsening symptoms resistant to treatment without previous investigation for IBD.

“So I suppose on anybody young under 45, anybody who had any symptoms of inflammatory bowel disease, so frequent bowel movements, any blood, any mucus, abdo pain with it, a strong family history with some of the symptoms, I’d have quite a low threshold for using it [FC test].” (GP04)

Referral was mentioned as an eligibility criterion for testing by several GPs. Some GPs initiated testing in patients not considered for referral, and referred rather than tested those who were believed to need an urgent referral. Another GP stated that they would test even if they were referring anyhow.

“... if somebody has the criteria that would justify doing the test then I think I would do it. (...) and even if I was going to refer them anyway it’s still useful information for the consultant.” (GP01)

5.5.2.3.3 Context

GPs considered the FC test within the context of information collated through history taking, examination and testing. Yet, GPs stressed that FC testing was complementary and supported rather than dictated decisions. Therefore, situations occurred where the test had little impact on the diagnosis or management decision.

“So, it wouldn’t be a case of just doing Calprotectin and if it’s positive sending the patient forward or not. So, I don’t think it has a huge amount of weight because, as I said, a good 60%, 70% of diagnoses are made from history and then you just supplement it with your investigations and examinations, so it’s only adjunct to the rest of what we’re doing. So, I don’t

think it's black and white, be all and end all of a diagnosis, that's the bottom line. It's useful but, yeah." (GP14)

5.5.2.3.4 Summarising the place of FC testing

My findings revealed that the severity of presenting symptoms determined whether FC testing was considered appropriate as part of the patient assessment and whether the test was used as a baseline or second line test. Eligibility was frequently discussed in terms of patient age and symptoms stipulated in local and national guidance providing little insight into who GPs really test. GPs described FC testing as complementary to their patient assessment and may test prior to referral, instead of referral or concurrently with referral.

5.5.2.4 Requirements for testing

"Well I mean obviously if it is recommended by NICE that's a good thing and it's been endorsed by a local consultant and it's been incorporated into a formal pathway." (GP01)

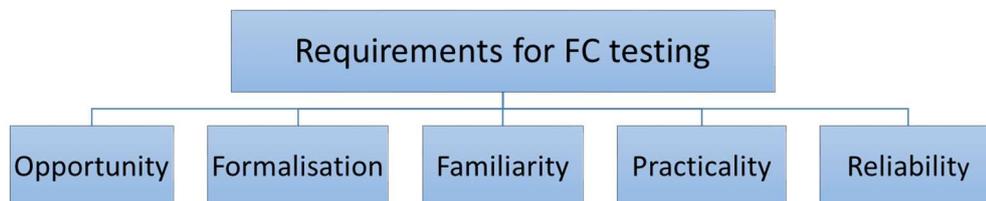


Figure 5.11 Requirements for FC testing according to GPs

My analysis suggested that availability of a test is instrumental for its use as it can create **opportunity**, but does not automatically result in uptake of testing into clinical practice. Further requirements (Figure 5.11) may need to be met. When considering a new test, GPs tended to look for local and national endorsement of the test for instance in way of **formalisation** in national guidelines or incorporation into local pathways. Other important factors that can promote test use were described as **familiarity** with the test, **practicality** of using the test and the test's **reliability** which influenced the GP's judgment of the usefulness of the test for patient management.

5.5.2.4.1 Opportunity

Based on my data, creating opportunity for testing required 1) that the test was available to the GP and 2) that patients with appropriate symptoms presented to general practice, which is not given in rare conditions like IBD. With regards to opportunity, GPs discussed the lack of situations that may prompt the consideration of testing, and that national guidelines were insufficient in making tests available locally.

“(...) it's aspirational what's in NICE and actually it doesn't happen for whatever reason if things aren't available or commissioned.” (GP05)

Once GPs learned about the availability of FC testing, they expressed a desire for equal opportunity.

“So if everybody else has got it, we want it as well. [Laughs]. Yes.” (GP03)

5.5.2.4.2 Formalisation

Following availability, formalisation of testing can aid the acceptance of a new test by GPs, which is required for its uptake into clinical practice. A test was accepted when it was formally embedded in a local guideline, recommended by NICE or stipulated on checklists for referrals. GPs expressed that they found it helpful if they were provided with a formal pathway explaining criteria for testing and steps to take following testing. GPs also looked for endorsement of tests by local consultants and opinion leaders when considering a new test which suggests that test use is motivated by what is expected of GPs and by what their colleagues do, particularly if they are more senior or more specialised.

“I think perhaps if the secondary care consultants were telling us that they would very much like us to be doing that test in primary care then great, (...). Perhaps advice from the local CCG in their guidelines (...). And I guess on a personal level if I am being more aware from my colleagues that it is being used and them having examples of how they find it useful.” (GP15)

5.5.2.4.3 Familiarity

Familiarity with FC testing was an important requirement for testing. This was thought to include awareness of the availability of the test, a feeling of competency

through sufficient training and information provided or acquired at the time of introduction, and personal experience with testing over time.

“So I think with any new thing it’s when you get used to doing it especially if you can see there is a benefit. (...) So you know. It’ll be sort of getting experience. Getting happy with it. And also knowing whether the gastroenterologists would pay attention to it.” (GP03)

GPs indicated that it was important to have the test present in the list of options when considering the management of patients and that the test can be done routinely without laborious and time consuming searches for information using up precious consultation time. GPs reflected on how they learn about new tests and reported that they would appreciate standardised information on how to use the test.

“I don’t really know how we get the information on the new tests, you’ve made me think about it, and it would be nice to have a good...like if a new test was coming, you would be told what are you testing, who are you testing, how it’s done exactly, (...) and what we’d worry about and any pros or cons to patients, any side-effects to them or problems that we might encounter doing it, it would be quite nice to have it standardised, but I don’t know, we just seem to get random emails and then teaching about it.” (GP04)

5.5.2.4.4 Practicality

As a further requirement for testing GPs considered the practicalities of FC testing. In this context, GPs discussed the cost of testing and the logistics of testing including turnaround times, knowledge of how to test and the ease of using the test. GPs also mentioned the acceptability of FC testing as well as convenience to patients as an influence on whether they would consider testing.

“And then also you have to think about the practicality of it. Patients don’t always like to deal with stools and things like that, they’d much rather have a blood test or something like that. So sometimes it’s just not practical for them, you just don’t like doing it.” (GP19)

5.5.2.4.5 Reliability

My analysis suggested that GPs need to know something about the test's accuracy before they would consider using it. This was thought to enable them to judge how useful the test would be for their decision making, what question it could help them with and how much weight they could place on a positive or negative test result. GPs also pointed out the importance of the timeliness of the results, knowing advantages and disadvantages of testing and whether there were any side effects for the patient.

“So I think really just having it clear in our heads why we are doing it and perhaps seeing over time that if we did start to use it more often, over time whether it actually is helping us distinguish patients and helping us in our referral pathways really. If it is still a little bit hazy and unclear, then perhaps it is not a valuable test.” (GP15)

5.5.2.4.6 Summarising requirements for testing

According to my data, national guidelines may be insufficient in making tests available locally. Furthermore, availability of FC testing may not be sufficient for its uptake into clinical practice. Several requirements needed to be met. First and foremost, GPs needed to be given opportunity for testing. The low prevalence of IBD meant that some GPs reported limited opportunity and familiarity with FC testing. Requirements, which GPs thought would encourage use of a new test, included the test's practicality and reliability, its formalisation into pathways and guidelines, and information on how to use the test.

5.5.2.5 Summarising FC testing within the referral decision

My findings suggest that outcomes of FC testing may have limited direct impact on the referral decision because a test result is not considered in isolation, and it is unlikely to override prior assessments from history and examinations. However, FC testing can shape the referral decision indirectly. FC testing has the potential to decrease referral rates as it may provide confidence in management decisions and reassurance for the patient and the GP. However, GPs acknowledged that FC testing is not truly diagnostic and only provides additional information. Therefore, a negative FC test is unlikely to automatically rule out a referral. FC testing can also result in uncertainty if not in line with prior belief and in challenges if it does not meet patient expectations. The outcomes of FC testing depended on who GPs test and

why. GPs tested for three main reasons: diagnostic uncertainty, reassurance and medico-legal protection. However, the role and place of FC testing within the referral decision varied in the GP's responses, which included whether a referral was considered without testing, concurrently with testing or following testing. Thus, I expect the impact of FC testing on referral rates to vary. My analysis suggests that in order to promote the positive outcomes of FC testing in clinical practice, GPs need to be provided with more and better information on the test accuracy and with guidance on who to test and how to use the test results. Referral decisions may benefit from FC testing if the appropriate use of FC testing can sufficiently increase confidence in patient management and reduce clinical uncertainty.

5.6 Discussion

5.6.1 MAIN EMPIRICAL FINDINGS ON REFERRAL DECISIONS

Referral decisions appear to be complex, individual and situational and were influenced by a number of factors. Based on my analysis I summarised these as clinical need, the doctor-patient interaction, GP factors and the health care system. The influences operate bidirectionally and I could not classify them as either motivating or discouraging a referral. Rather, influences could be seen as sliding scales, interrelated and working together to different degrees to determine the outcome of individual referral decisions. According to the GPs' accounts, the referral decision did not follow a pathway but resembled a jigsaw puzzle of information which was never fully complete. It explains how different responses to different challenges result in individual GP referral thresholds, a concept first suggested by Cummins et al. (1981).¹⁹³ Generally, GPs considered referral if: they thought symptoms indicated a referral; there was clinical uncertainty for instance because symptoms were unresponsive to treatment; guidelines prescribed a referral; a patient communicated the preference for a referral; or the GP's risk threshold of missing something serious had been reached. Referral decisions were based on patient assessment informed by history, examinations and investigations. If FC testing was initiated to inform the patient assessment the result was unlikely to override prior assessments from history and examinations and it was unlikely to be interpreted detached from the testing context. FC testing had, therefore, little direct influence on the referral decision but affected the main referral influences in various ways. The direction of the effect was difficult to predict and would depend on the reason for testing, the test result, prior belief, the patient, the doctor and their interaction. Every patient presented a unique combination of factors for a GP and I

could not identify any clearly defined categories in which patient encounters could be placed that would predict the outcome of a referral decision.

I discuss the main study findings in more detail under the following headings: challenges in primary care and types of GPs.

5.6.1.1 Challenges in primary care

A major challenge in primary care is clinical uncertainty. Serious disease was reported to be rare but GPs felt they were expected not to miss it. The risk of missing something serious and its consequences including patient harm and litigation, forced GPs to practice more defensive medicine which has been reported to be a result of unsupportive colleagues and management in cases of adverse outcomes and complaints.¹⁹⁴ Therefore, “When in doubt it’s good to refer.” (GP08) was motivated by both, the patients’ well-being and the GPs’ own protection. However, the pressure not to miss cancer and to refer suspected cancers in a timely manner was conflicted by the pressure to keep referrals low.

Inappropriate referrals were discussed by several GPs as well as the pressure from CCGs to contain and to justify referrals. GPs described that CCGs scrutinise and league table practice referrals and name and shame high referring practices. This is of concern as overall referral rates cannot be used as a simple proxy for referral quality.¹⁷² Simply reducing referral rates is likely to reduce appropriate and inappropriate referrals and high referring practices do not necessarily have high proportion of inappropriate referrals.¹⁷⁴ One of the criteria the CCGs use to define inappropriate referrals is discharge at first outpatient appointment (GP10). This would condemn many referrals for reassurance where the appropriateness clearly depended on who judges referrals. Taking the perspective of specialists, 13% of referrals were assessed as inappropriate with good agreement between the specialists and an independent GP.¹⁹⁵ Elwyn and Stott (1994)¹⁹⁶ concluded that 35% of 168 referrals could have been managed in primary care while a study taking the patients’ perspective reported unsurprisingly that 95% of patients rated their referral as necessary and worthwhile.¹⁸⁴ This is because referrals that are not clearly necessary on clinical grounds may still be justified in terms of providing reassurance to the GP and/or patient. Appropriateness is, therefore, context specific and should be defined in terms of necessity, timeliness, destination and process which will vary for different stakeholders.¹⁷² Moreover, removal of all inappropriate referrals judged by receiving specialists in one study only reduced a 2.5-fold

variation in referral rates to a 2.1-fold variation.¹⁹⁷ Therefore, the problem in referral variation may not be over- but under-referral.¹⁷⁴ Appropriateness of referrals is clearly difficult to measure and define and needs to consider patient benefit and clinical outcomes.¹⁷⁴ However, a study relating clinical outcome (i.e. stage of cancer at referral) to variations in GP referral rates found no adverse outcomes for breast and bowel cancer patients associated with low GP referral rates.¹²¹ As the optimal referral rate for GPs and GP practices is unknown, it is alarming that they are used to manage referrals downwards.

A further challenge in primary care is time. Time was discussed in different contexts of the referral decision. On the one hand, GPs discussed the short consultation time as pressurising and unsupportive for patient management which may lead to hasty referrals. On the other hand, time was described as a “great diagnostician” (GP16) which is a difficult concept for patients who want a “quick fix”. GPs were also concerned with timely referrals and long waiting times for routine secondary care appointments. While GPs claimed that referral decisions are (purely) clinical decisions they acknowledged that long waiting times affected their referral decisions.

Another challenge mentioned was meeting referral targets as well as expectations to offer patient choice. These expectations are “at odds” (GP05). During the 2000s policies were introduced that committed the NHS to provide more choice to patients over time, date and destination of their referral.¹⁹⁸ As a result an electronic referral system was implemented as a top–down policy led by the Department of Health. The main aim was to reduce waiting times rather than referral rates. However, little research evidence is available on the effectiveness of the system while surveys reported that GPs feel negative about the system due to its inflexibility, unreliability and inability to refer to a named consultant.¹⁹⁸ As a result, uptake was slow and my interviews confirmed that some GPs did not like to be forced to use the system and still resorted back to ‘paper’ referrals. Dixon et al.(2010)¹⁹⁸ concluded that it is largely dependent on the GP whether a patient is offered choice at the point of referral.

A different attitude towards patient choice was also discussed by GPs in the context of referral requests. GPs were in a particularly difficult position when patients could not be reassured in primary care and preferred to see a specialist. GPs may refer despite their knowledge that a referral is not going to change or add to the patient’s

management. Depending on the GP's personality and the type of patient this might be experienced as a personal failure in patient management or as patient choice. This suggests that GPs may respond differently to patient pressure depending on how they perceive it. This supports the finding that different perceptions of pressure by GPs can describe some of the variation in referral rates among general practitioners.¹⁹⁹

5.6.1.2 Types of GPs

Differences in GPs also emerged in their perceived gate-keeping role and their definition of a wasted referral. I explored whether there are certain types of GPs with certain ideas about referral associated with the responses to the ideas of wasted referral, patient pressure and perceived gate keeping role. However, this revealed no pattern to sufficiently support the distinction of "gatekeepers" (referral is a last resort and needs justifying, if not justified it is a waste) and "advocates" (acting on patient's behalf, every referral has got a reason) into which all GPs would fit. GP11 would qualify as an advocate. The GP felt very strongly that GPs are no longer gate keepers, it is the patient's right to ask for a referral and that no referral is wasted. In this case it seemed the GP's personality and attitude towards referral in the current NHS system that prompted their referral decision. This can be contrasted with GP05. This GP identified themselves as a gatekeeper, described referrals in response to patient's wishes as "giving in" and felt the need to be able to justify referrals. Other GPs, however, showed aspects of both. This finding agrees with that of an interview study of GPs in the management of patients with minor mental illness.¹⁸³ This study characterised GPs according to two distinct referral strategies. These were containment (minor mental illness is part of GP's remit, GP's role is to manage not refer) and conduit (GP's role is to diagnose, triage and refer). The authors further concluded that GPs do not simply fall into one or the other category but that GPs use both strategies and that there was a continuum of behaviour.

Variation in GPs attitudes and characteristics was not only reflected in the content of their responses but also by the different language they chose. Using the topic of returning patients as an example, GP13 talked neutrally about patients that "...may keep coming back regularly... And sometimes it does get to a stage where you think you may need to refer these patients for reassurance." The language of GP06 was much more emotive describing a similar situation: "Patients will vote with their feet in a way and just keep coming back. ...if they come back after two days, or after four days, and after six days...eventually you do get worn down...you often feel you

have got no alternative than refer them on.” While the language used suggested to me that GPs have different personalities and attitudes towards referral, management and testing, the data again did not support grouping GPs into different categories. Thorsen et al. (2016)²⁰⁰ used findings from a survey with Norwegian GPs to describe two different ways in which GPs think and work when they refer. The two characteristics were “confidence” and “uncertainty”. However, they did not confirm the findings with in depth qualitative methods and I believe this may be too simplistic considering my interview data.

Bailey et al. 1994²⁰¹ suggested the definition of disease-centred and patient-centred for high and low referrers, respectively following their quantitative content analysis of GP interview data. However, they could not confirm this with their qualitative analysis but found that interpersonal aspects of the consultation featured more prominently in the responses of low referrers and more uncertainty in high referrers. While my data prevented me from drawing conclusions about high and low referrers, I would argue that GPs can be both, disease-centred and patient-centred based on what they reported about different patients and the clinical problems they presented with.

These examples seem to confirm the difficulty of categorising GPs according to their referral behaviour. Even though I saw considerable variation in the GP responses I think the responses to the interview questions were situational and only in part due to differing GP characteristics. GPs chose their own stories they wanted to share and all reflected on different patient encounters which hampered a direct comparison. The specific patient story the GPs were describing will have influenced their answer and they may have answered some questions differently if they had described a different patient situation. Anchoring interview questions in patient vignettes might be an alternative approach to reveal differences in GPs’ characteristics.¹⁸²

5.6.2 THEORETICAL FINDINGS IN THE CONTEXT OF PUBLISHED MODELS OF REFERRAL DECISIONS

The main referral influences in my study were in agreement with the findings from studies dating back to the 1980s and 1990s.^{175, 176, 178} Studies reported that clinical factors, GP factors and structural factors (the health care system) influence referral decisions. My study confirmed that these referral influences are still valid and are applicable to referral decisions particularly involving patients with abdominal

symptoms. However, my original contributions to referral theory included that some of the influencing elements within these factors might be shifting and changing and that referral models may be outdated and insufficiently reflective of the complex referral decision context.

In 1983 Dowie¹⁷⁸ published the first study which investigated the process of the referral decision. According to her study the influences on a referral decision can be grouped under three headings: professional attributes, personal style and knowledge of the health care system. Dowie identified the reasons for variations in referral decisions to be the GPs' cognitive processes. These included differing confidence in the GPs' clinical judgment and differing awareness of the base rate (prior) probabilities of serious conditions. She also suggested that GPs have different levels of medical knowledge and that they work towards sustaining the esteem of consultant colleagues. My analysis revealed that there are additional non-clinical factors, including the GPs' perceived role in the health care system, their differing perception of patient pressure and their differing responses to the pressure from the health care system, which are equally important. This agrees with the findings by Newton et al. (1991)¹⁷⁶ who argued that many referral factors have little to do with GPs' reasoning processes. Furthermore, my data indicated that the changing health care system provides additional direct influences in terms of guidelines and CCG referral charts, which was not mentioned in previous studies. Another difference appears to be a shift from a GP focused doctor-patient interaction to patient centred care. GPs in my interviews stressed the importance of patient choice, negotiations and shared decision making. In contrast, earlier studies described the interaction with patients as a relationship of the GP with the patient which seemed more GP centred. The doctor-patient interaction theme which emerged from my data reflected this change and captured a wider range of interactional styles.

Recognising non-clinical factors and the doctor-patient interaction as important influences on referral decisions was confirmed by a systematic review on interventions to manage referrals.²⁰² The review found that there was strong evidence that interventions affecting GPs attitudes and beliefs as well as the doctor-patient-interaction can improve referral, while interventions that affect the GPs' knowledge or referral behaviour have a weaker or conflicting evidence base in their impact on referral.

In essence, my study supported the findings of factors that influence referral decisions which may explain some of the variations in referral rates, but they cannot explain the working of the influences and processes within individual GPs' referral decision making. Two main findings emerged from my data which may point towards the working of the influences. Firstly, influencing factors interact in the referral decision making process, which was echoed by the findings by Newton et al. (1991).¹⁷⁶ Secondly, referral decisions are not linear. This is in contrast with the individual referral decision model by Dowie (1983).¹⁷⁸

Dowie (1983) took the evidence on referral influences to the next level and modelled the GPs' individual referral decision based on conflict theory.¹⁷⁸ Janis and Mann's conflict model (1977), summarised in Dowie (1983),¹⁷⁸ views humans as reluctant decision makers and describes how people cope with decisional conflicts. They argue that decisions of real consequence generate psychological stress for individuals and that individuals make use of five mechanisms to cope with or avoid that stress. Dowie was the first to try and explain how GPs make the decision to refer for diagnosis by using this model and explaining decision making under conditions of uncertainty. The model consists of four main yes/no questions that GPs deal with during the decision process following the assessment of the patient:¹⁷⁸

1. Are the risks to the patient serious if I don't refer now?
2. Are there risks to my esteem serious if I refer now?
3. Is it realistic to hope to find a better solution?
4. But is there sufficient time to search further and deliberate?

Some cases will cause little conflict and stress because the referral is clear cut, or the situation is well within the GPs management capabilities. If the case is more complex and the situation involves risks, the doctor's stress intensifies. They may postpone the decision (defensive avoidance) if the risk is small or shift responsibility by referring if the risk is serious. They may refer urgently without sufficiently evaluating the case because they are anxious that there is not enough time to investigate. If the GP feels there is time to gather more information GPs may continue a thorough search and evaluation of evidence. Differences in GPs level of anxiety, self-confidence and insecurity may explain different use of these coping mechanisms.¹⁷⁸ Wilkin and Smith (1987) developed the model to be more inclusive of other referral reasons.¹⁷⁵

The models portray the referral decision process as a progression through a sequence of questions requiring yes/no answers. However, considering my findings, I would question whether complex real-world decisions are made using such a linear approach, which does not consider the patient influence. Furthermore, I feel that the model fails to consider the context in which referral decisions are made, the health care system and the GP as a person. These are important factors as my interview study revealed. The idea of heuristics and biases²⁰³ supports the view that referral decisions are not made in such a reasoned manner. Kahnemann (2012) argued that it is human nature to adopt shortcuts (heuristics) in decision making, such as intuition, which may result in errors (biases) in cases of incorrect assessment of prior probabilities and failure to consider all information and relevant alternatives.²⁰³ GPs in my study reported that they sometimes acted on a clinical hunch while at other times they weighed up probabilities. This fits with the cognitive continuum theory by Hammond 1981 which places clinical decisions on a continuum between pure intuition and pure analysis.²⁰⁴ Others agreed that easy cases, judged on the level of experience and knowledge of the specific context, will result in more intuitive decision making, while complex cases will require analytical hypothesis formation.^{185, 205} There appears to be a need for more research into combining referral decision theory and clinical decision theory to explain complex referral decisions.

Furthermore, the models of referral and clinical decision making are GP centred, where the GP is the decision maker and the patient is the object. They do not consider what GPs in my study described as joint decision making or situations in which patients oppose a management decision. This paternalistic decision making model,²⁰⁶ appears outdated considering the responses to my interviews which reflect that today's health care system emphasises shared decision making.²⁰⁶ Shared decision making requires strategies to create an environment in which patients feel comfortable to disclose information necessary in arriving at a decision, to avoid situations that may break up the doctor-patient relationship and to reach management decisions that are acceptable to patients.²⁰⁶ Elliott (2010)²⁰⁷ termed this 'mutual intacting' (both patient and practitioner are working to keep the relationship intact). She argues that the psycho-social issues described by 'mutual intacting' need to be integrated with existing clinical decision theory. My findings agree with her view as the GPs were genuinely concerned about their relationship

with the patient and stressed the importance of negotiation and compromise in reaching an acceptable management decision.

Overall, we still lack a model that explains the complex decision making process of GPs that is applicable to all referrals, considers the decision making context as well as the concepts of patient choice and shared decision making.

5.6.3 FC TESTING AS A REFERRAL DECISION AID

My original contribution to referral theory development included the addition of a testing question to the referral question. FC testing is one recent example of making available a test in primary care as a referral decision aid. Its impact on the referral decision is questionable considering my findings in sections 3.5.4.2 and 4.3.4.2.8, which showed that referral decisions do not always follow the results of FC testing. The reasons behind this became apparent in the GP interviews. Therefore, this study adds a vital perspective on Chapters 3 and 4 as it explains why such a great proportion of patients with negative FC tests are referred.

GPs appear to need stronger evidence not to refer than to refer. While a positive test result or concerns over presenting symptoms may have been sufficient to warrant a referral, a negative FC test result alone was not sufficient to prevent a referral as GPs were always left with the question “What if...?”. The problem with FC testing is that it excludes IBD reasonably well but not all other organic conditions. Therefore, GPs still need to refer. One GP (GP11) pointed out that neither the evidence nor the guidelines explicitly say not to refer if FC is negative. Therefore, the risk of missing disease and litigation remains. This may explain the apparently inappropriate referrals of patients who tested negative. Considering the discussion of inappropriate referrals in section 5.6.1.1, I would argue that in cases of referrals of FC negative patients, the test failed to provide sufficient objective reassurance to prevent a referral.

This leads to the question who GPs test and for what reason. The interviews revealed that GPs tested for three main reasons: diagnostic uncertainty, reassurance and medico-legal protection. However, published FC pathways only consider diagnostic uncertainty as an indication for testing.^{58, 118} Furthermore, the starting point of pathways is diagnostic uncertainty of IBD and IBS. This is a secondary care perspective and fails to acknowledge the broader spectrum of symptoms and disease in primary care and the needs and motivations of GPs. FC

pathways should consider this broader role that GPs ascribed to FC testing and provide clearer guidance on who is eligible for testing.

When considering the question of how the FC test might fit in the referral decision pathway I need to rephrase the question, as the interviews showed that the referral decision is not linear. Furthermore, FC testing is not truly diagnostic and only provides additional information. I, therefore, consider the FC test as an additional piece to the jigsaw of information to inform the referral decision. My data suggest that its place depends on the severity of presenting symptoms and the reason for testing. This determined whether FC testing was considered appropriate as part of the patient assessment, whether the test was used as a second line or baseline test, and whether a referral was considered without testing, concurrently with testing or following testing. FC testing can influence a decision towards or against a referral because I showed that the many outcomes can affect the decision in different ways.

The overall direction and size of impact of FC testing on referral cannot be predicted from the interviews. Overall, the hope was that it cuts down on periods of observation, confers reassurance and ultimately reduces referrals. However, some GPs remained unconvinced that the FC test will have a big impact on referrals.

Considering FC testing within the referral models, I found that FC testing cannot easily be slotted into the referral process to predict its impact on the referral decision. In Wilkin and Smith's model¹⁷⁵ FC testing could provide additional knowledge for the referral decision. However, this scenario only applies to FC tests for diagnosis rather than for reassurance or protection. The model would need a question on the GPs' and patients' confidence in the proposed management strategy to improve inclusiveness.

I also mapped the outcomes of FC testing identified in my study to the immediate effects of referral interventions identified in the review by Blank et al. (2015).²⁰² This suggested that FC testing has the potential to affect the GPs' attitudes and beliefs and the doctor-patient interaction. This is in line with my findings. Both aspects are related to a positive impact on referrals.²⁰² However, this could imply an increased satisfaction or provision of adequate referral information, while the evidence of a positive impact on the referral rate by addressing GPs' attitudes and the doctor-patient interaction was weak.²⁰² Therefore, the model cannot be used to determine the possible impact of FC testing on referral rates.

Overall, GPs saw their main role as taking a good history in order to work out what the patient needs and wants. Testing was described as secondary because the majority of patients generally do not need tests. Furthermore, GPs are generally good at referral decisions on clinical findings and in some instances need to move away from using investigations to confirm clinical suspicion.²⁰⁸ Tests can create worry and false results which cannot be ignored. What motivates testing in primary care is that testing can happen sooner than testing in secondary care, it helps with the ownership of management and keeps patients out of hospital.

More research is needed to assess the impact of FC testing which takes into consideration the growing understanding of the process of referral decisions, which takes a whole system approach, including patient and clinician views, and which measures, both, outcomes of referrals and non-referrals.

5.6.4 STRENGTHS AND LIMITATIONS

5.6.4.1 Sampling and recruitment

The GPs interviewed in this study were recruited from a geographically diverse area stretching over seven different CCGs. This provided different perspectives from GPs with and without available FC testing, experiences with different hospitals, a range in waiting times and consultants. I succeeded in recruiting GPs with a wide range of backgrounds including those with and without experience with FC testing ensuring a wide range of perspectives was captured.

Recruitment continued until saturation was reached. This was possible because recruitment was iterative and interviewing and coding was carried out concurrently.

5.6.4.2 Study design and analysis

I chose interviews with individual GPs because I wanted to try and understand the GPs' individual decision making process of referral decisions focusing on individual reactions and individual experiences. I chose the GP perspective because GPs are the first contact for patients and play a major role in initiating the referral process. By talking to GPs I was able to gain insights into how they make a decision to refer and how they view patients and their influence in that process. However, I did not observe the referral decision process. I have to consider that what GPs said they do might not be what they actually do (but what they think they do or what they ought to

do), that GPs only talked about the things that they felt worth mentioning and that certain patient stories might not be reflective of the norm. Equally, I did not study the patient's perspective. As a result, the data I was able to capture might be incomplete, unbalanced and subjective. To clarify that the patient factors influencing a referral decision were perceived influences by the GP, I chose not to add a theme 'patient factors' but summarised the GPs' views on patients within the themes 'GP factors' and 'doctor-patient interaction'. However, the input of the PPI advisors into the project provided a patient perspective on referral decisions. An understanding of the whole referral process would require additional studies with patients and various secondary care providers.

GPs classified referrals for non-cancer conditions into urgent referrals and routine referrals. However, I was unable to elicit from the data how a GP distinguishes whether a patient's case was urgent or routine and to what extent such distinction was personal to individual GPs. This would have been of interest to further understand the different characteristics of GPs, but would not have generated more knowledge of GPs' referral decisions (refer or manage); only on the type of referral made.

5.6.5 VALIDATION OF STUDY FINDINGS

I shared a summary of my preliminary findings with all participants and invited GPs to comment in order to help me establish whether my account of their contributions is credible and trustworthy. I have already discussed in Chapter 2 that there has been much debate whether qualitative research can be validated in a similar way to quantitative research findings. I found that within the critical realist paradigm there are appropriate techniques available that can be used to assess the credibility of qualitative research findings.⁷⁹ Lincoln and Guba (1985)²⁰⁹ proposed that member check is the most crucial technique for establishing credibility. In line with the views of Angen (2000)⁷⁹ and Harbison et al. 2001²⁰⁴ I used the member check to establish the trustworthiness and confidence in my findings. 2/19 GPs responded and I considered their specific comments in the final write up. The feedback was very positive and the GPs generally agreed with the presentation of the complex factors influencing referral decisions. Both GPs commented on the referral in response to patient pressure, which one thought to be "unduly negative". However, both GPs agreed that these consultations exist but are uncommon as successful communication and negotiation usually promote clinically based decisions. I was pleased to receive the two responses. However, this meant that 17 GPs had either

nothing to comment or had no time to read my summary. More comments might have resulted in more disagreement. I interviewed 19 GPs from seven CCGs with varied patient experiences. I would, therefore, not expect everything in the write up to resonate with the personal experiences of each GP. The interviews aimed to explore the wide variation of influences and views rather than identify the most common situation. The feedback which included that the report clearly and effectively presents the findings adds to the confidence and trustworthiness of the study.

5.6.6 REFLECTIONS

As a researcher with a background in health technology assessment supporting national decision makers I had to relearn three main preconceptions before I could fully understand and make sense of the GPs' stories.

Firstly, I embarked on this study with a public health perspective. My questions were broad and my expectation was that the GPs' responses would be applicable to the general population of patients with abdominal symptoms, IBD or IBS. However, GPs are decision makers for individuals and are responsible for their health and well-being. They appeared to think more in terms of individual patient stories and their accounts were on individual patients. Some GPs found the questions too broad considering that every patient encounter is different. These GPs did not like general questions like "What are the advantages of testing?" and needed to know which test and in which context. Other GPs were a lot happier to reflect on their experience and to choose examples that would tell their story. It took a few interviews for me to fully appreciate this difference in perspectives and the potential impact. I kept my questions broad to give GPs the opportunity to respond as they felt best fitted but added the prompt to think of a specific example if they struggled with the question. This made me think that some responses might have been different if GPs had chosen a different patient story to talk about. I learnt from the interviews that clinical decisions are context specific and unique and that general pathways are often unhelpful in specific situations.

Secondly, I had to learn that referral pathways are not linear as I envisaged them. Once I stopped trying to fit the GP's responses into my mental pathway (patient assessment – testing - test result – diagnosis - referral decision) I started to understand how complex the decision making process is and that the referral decision process resembles a jigsaw puzzle with missing pieces.

Thirdly, my understanding of diagnosis was not necessarily compatible with that of GPs'. As a methodologist in test evaluation I frequently categorise patients into those with and without the target condition in order to calculate test accuracy metrics. This is quite final and should be followed by appropriate treatment. For GPs a diagnosis was often temporary, it could evolve and change. It was described as a working diagnosis that informs immediate management decisions which can include other tests. GPs often used referral and diagnosis interchangeably which seemed to confirm that a diagnosis is a management decision rather than a permanent label.

Perseverance and constant comparison during analysis enabled me to refocus my thinking and stay true to the data.

5.7 Summarising the main influences on referral decisions and the impact of FC testing on referral decisions

In this chapter I reported, interpreted and discussed the influences on GPs' referral decisions of patients with abdominal symptoms. Using framework methodology,⁷² I interviewed 19 GPs to explore their views and experiences related to referral of patients with chronic abdominal symptoms, to gain a deeper understanding of the influences on and thinking behind referral and reasons for testing. Four major themes emerged from the data which described the main influences on a referral decision. These four influences were clinical need, the doctor-patient interaction, GP factors and the health care system. The analysis revealed that referral decisions are not linear. They are complex, individual and situational because every patient consultation is unique and the influences are interdependent. While all four influences operate on every decision they can carry different weight in different situations.

My findings support those of published studies investigating referral influences mainly from the 1980s and 1990s. This suggests that the main referral influences are still valid and that the variability in GPs' attitudes and beliefs including their perceived role, their confidence and tolerance of risk, which may result in individual referral thresholds, are not condition specific. However, my original contribution includes that changes in the health care system resulted in additional direct influences in form of guidelines and referral charts which GPs reported can have a great impact on their decision making. Secondly, I identified an increased focus on

shared decision making and patient choice, which placed greater importance on the doctor-patient interaction as a referral influence. Thirdly, my data suggest, that existing referral models portray the referral decision as too linear, they are too GP centred and lack consideration of important non-clinical factors and of different referral contexts. New models should take a whole system approach and integrate psycho-social issues and clinical decision theory with referral decision theory to explain complex referral decisions.

Referral decisions can be made with and without testing. If a test is indicated the result provides additional information to inform the management decision which may involve the referral of the patient. A test result is not considered in isolation, however. It is regarded within a wealth of information from history and examinations. The influence a test result has on a management decision depends on a multitude of factors including the reason for testing, the test outcome, prior belief, available guidelines and patient expectations. Outcomes of testing, both positive and negative, may indirectly be supportive or unsupportive of the referral decision or have no impact on the referral decision at all. A negative test is not always sufficient to prevent a referral, because FC testing is only one piece of information in the patient assessment and cannot alleviate all uncertainties and risks. Referral of FC negative patients are likely to be due to preferences or uncertainties not addressed by FC testing. Whether these referrals are inappropriate depends on the perspective of those who judge the referral. The GPs' opinion that FC testing has reduced referrals still needs to be confirmed in studies. This study adds a vital perspective to Chapters 3 and 4 as it explains why such a high proportion of patients are referred following a negative FC test result.

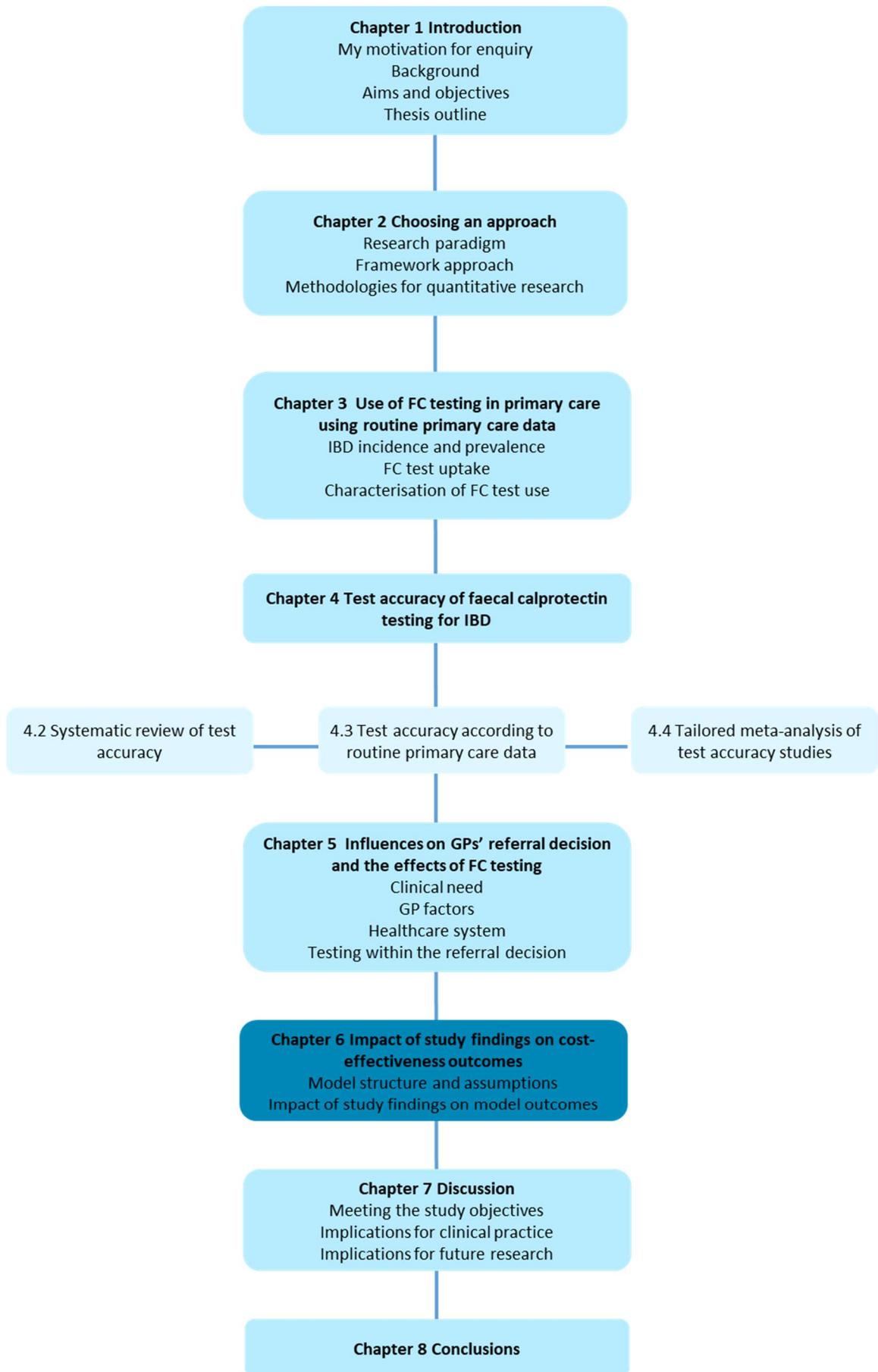
If a diagnostic test is to be used as an effective referral decision aid it needs to be able to

- influence the individual GP's beliefs and attitudes,
- support the doctor-patient relationship,
- provide confidence in the management decision and/or
- have a referral pathway to trusted services available.

As these are unlikely to be a given for each referral decision, variation in referral rates following the introduction of FC testing is likely. The complexity of referral decisions illustrates that we require a whole systems approach for the development

and evaluation of referral decision aids including understanding how a test is going to be used in clinical practice, the impact on referral and patient outcomes following a referral decision or the decision not to refer. Considering the FC test outcomes within existing referral models to predict the impact of testing on referral decisions was not successful. This was because models were not sufficiently inclusive to allow for different reasons for testing, evidence on the impact of referral interventions on the referral rate is to this end inconclusive, and possibly because testing outcomes and their effects on referral influences are too complex, as my data showed, to predict referral decisions.

The finding that FC test negative referrals may not be inappropriate and the view that under-referral rather than over-referral may be the issue that leads to the variations in referral rates that have been observed for decades suggests that it may be inappropriate to pressure GPs to reduce their referrals. This means that the number of test negative referrals should be considered in the health economic evaluation of FC testing in primary care. An adjustment of cost-effectiveness claims may be needed. This forms part of my discussion of the cost-effectiveness model in Chapter 6.



Chapter 6 Impact of study findings on cost-effectiveness outcomes in the NICE assessment

6.1 Chapter overview

The aim of my thesis has been to establish primary care evidence for FC test use (Chapter 3) and FC test accuracy (Chapter 4) in order to discuss the validity of health economic estimates in the NICE assessment for the primary care context which led to the national approval of FC testing in primary care.

In this chapter I discuss the potential impact that my study outcomes may have on the cost-effectiveness of FC testing given the study limitations. The discussion includes a description of the NICE model, the model inputs and assumptions, the direction of likely change of the cost-effectiveness and the certainty I am able to place on these predictions considering the study limitations.

Initially, I had planned to update the model inputs with the outputs from my studies and revise the cost-effectiveness estimates for FC testing in primary care. However, this was unfeasible for the following reasons. Firstly, the model considered FC testing for IBD versus IBS, which is not applicable for routine primary care. Adjusting the model for the differential IBD versus non-IBD would require adequate knowledge of the various other non-IBD inflammatory diagnoses. It would also require a decision on how to value these conditions in terms of costs and utilities depending on whether they follow the IBD or non-IBD pathway. However, collecting detailed information on non-IBD cases was outside the scope of my study objectives. Secondly, as discussed in Chapters 3 and 4, my outputs of referral and test accuracy are based on a number of assumptions and incomplete data due to the nature of routine electronic health records data. I, therefore, feel that producing a single estimate of cost-effectiveness based on these outcomes would not be appropriate. It would not be reasonable to attempt to update the current cost-effectiveness estimate, potentially suggesting a revisit of the NICE approval on the basis of such figure.

6.2 Health economic model

6.2.1 MODEL STRUCTURE AND ASSUMPTIONS

The health economic model by Waugh et al. (2013)⁵⁸ assessed the cost-effectiveness of faecal calprotectin testing compared with GP assessment alone for the differentiation of IBD and IBS in primary care in the adult population.

The model consisted of two parts. The first part was a decision tree which categorises the cohort correctly or incorrectly into IBD and IBS using sensitivity and specificity estimates from the secondary care literature as probabilities. This part of the model was coupled with a Markov model representing the treatment pathways for the different conditions. Patients follow the respective treatment pathway for Crohn's disease, ulcerative colitis and IBS based on their classification by the decision tree. The linked-evidence decision analysis combined the outcomes of diagnostic strategies with patient management to allow the estimation of clinical outcomes, costs and quality adjusted life years (QALYs). Quality of life (QoL) impact was dominated by patients with false negative results not being correctly treated for IBD and, therefore, not entering remission. The time spent with a false negative result was of importance here. The second QoL impact was QoL decrements due to adverse events from colonoscopy.

Costs considered included costs for FC testing, colonoscopy, drugs, administration of drugs and outpatient appointments with a consultant or nurse. The model estimated long term outcomes of costs and consequences of testing and treatment over a ten year time horizon. The model adopted the NICE reference case perspective of patient benefits and NHS and Prescribed Specialised Services (PSS) costs and discounted costs and benefits at an annual 3.5%.

Model assumptions for the base case for primary care included that:

- Sensitivity and specificity are transferable from secondary care studies
- Base case prevalence of IBD for primary care is 6.3% in the eligible population
- Only patients with a positive FC test are referred to secondary care
- All patients referred to secondary care receive colonoscopy
- Colonoscopy is 100% specific meaning that the overall test pathway has no false positive outcomes

- Delay between referral and colonoscopy is four weeks and the time to retesting among those testing negative but not responding to IBS management is 12 weeks
- Patients with a negative colonoscopy result go back to primary care for IBS management

6.2.1 MODEL BASE CASE PARAMETERS AND OUTCOMES

The base case was derived from consideration of the cost-effectiveness of GP assessment without FC testing compared to FC testing using a 50µg/g threshold. Sensitivity and specificity to detect IBD by GP assessment alone was taken from Otten et al. (2008).¹³³ Summary estimates of sensitivity and specificity for FC testing were derived from five studies of FC testing in the diagnosis of IBD versus IBS from secondary care from the systematic review by Waugh et al. (2013)⁵⁸ (Table 6.1). Patients receiving FC testing were those considered for referral by the GP if no FC testing is available (25% of patients with symptoms).

Table 6.1 Estimates of sensitivity and specificity considered in the cost-utility model and estimates based on my THIN data study (reported in section 4.3)

Scenario	Sensitivity (%)	Specificity (%)
GP assessment only (base case)	100.0	78.8
GP assessment only (wider population)	94.1	89.7
FC testing for IBD vs IBS Waugh et al. (2013)	93.0	94.0
THIN primary analysis	92.9	61.5

FC faecal calprotectin, GP general practitioner, IBD inflammatory bowel disease, IBS irritable bowel syndrome, THIN the Health Improvement Network

Compared to GP assessment alone, FC testing had lower sensitivity (Table 6.1). This resulted in a slightly greater decrement in QoL due to more people with false negative results spending time without treatment for IBD and, therefore, not achieving remission. Specificity of FC testing was higher than for no testing resulting in lower numbers of false positives (Table 6.1). This reduced the decrement in QoL due to adverse events from colonoscopy and more than offset the initial costs of FC testing because of fewer colonoscopies (Table 6.2). GP assessment without FC testing was never cost effective when compared with calprotectin testing due to an overall small QALY gain coupled with cost savings. Because of the low prevalence and high sensitivity of FC testing the number of false negative tests was small and, therefore, the QoL decrement was small. The cost effectiveness was driven by the

cost of testing, including the FC test and colonoscopy, and possibly also by the QoL decrement due to adverse events from colonoscopy.

Table 6.2 Results from base case and scenario analysis considering a wider testing population

	QALYs	Costs (£)
GP base case	6.2285	3297
FC testing base case	6.2291	3215
GP wider population	6.1952	2978
FC testing wider population	6.1955	2971

QALYs quality adjusted life years, GP general practitioner, FC faecal calprotectin

6.2.3 SENSITIVITY AND SCENARIO ANALYSES

6.2.3.1 Population considered for FC testing

The model explored the possibility that FC testing might be considered more widely than only in those considered for referral (the base case assumption). A sensitivity analysis was used to assess testing twice as many patients as would be considered for referral (50% rather than 25% of patients with symptoms) resulting in a lower prevalence of IBD in the tested population (Tables 6.3 and 6.4). Testing in a wider patient group resulted in a smaller number of false negative results but also in a greater number of false positive results. Cost savings were markedly reduced coupled with small QALY gains still rendering FC testing narrowly dominant over GP assessment alone (Table 6.2).

Table 6.3 Disease prevalence and proportions of true positives, false negatives, true negatives and false positives in detecting IBD by GP assessment versus FC testing for base case assumptions and the sensitivity analysis of a wider testing population

	Base case		Wider population	
	GP assessment	FC testing	GP assessment	FC testing
IBD tested, %	6.3	6.3	3.4	3.4
TP, %	6.3	5.9	3.2	3.2
FN, %	0.0	0.3	0.2	0.2
IBS tested, %	93.7	93.7	96.6	96.6
TN, %	73.9	88.1	86.7	90.8
FP, %	19.8	5.6	9.9	5.8

FC faecal calprotectin, FN false negatives, FP false positives, GP general practitioner, IBD inflammatory bowel disease, IBS irritable bowel syndrome, TN true negatives, TP true positives

6.2.3.2 Prevalence of IBD

Prevalence was varied in sensitivity analyses from 5% to 25% in 5% increments. A higher prevalence resulted in a greater number of false negative results. Therefore, a high sensitivity would be more important in populations with high prevalence.

6.2.3.3 Not all referred patients receive colonoscopy

In the base case analysis all referred patients received colonoscopy. This was due to a lack of information on possible referrals without resulting colonoscopy in some patients. In this scenario analysis, different specificities (25%, 50%, 75% and 95%) were explored of the process of gastroenterological assessment to exclude IBD at an assumed sensitivity of 100%. At a specificity of 95%, FC testing could still avoid the cost of unnecessary referrals but was no longer instrumental in avoiding the cost of unnecessary colonoscopies. Cost savings dwindled as the specificity of gastroenterologist assessment increased. In the scenario where a wider patient population was tested, the FC testing strategy was more costly than GP testing without FC testing.

6.2.3.4 FC cut-off

Based on test performance measures looking at different cut-offs, the 50µg/g cut-off was dominated by the 100µg/g cut-off because of the poorer specificity at the lower cut-off.

6.2.3.5 Referral rate

The referral rate was not considered in sensitivity analyses.

6.3 Potential impact of findings from Chapters 3, 4, and 5 on cost-effectiveness findings

6.3.1 DISCUSSING MODEL OUTCOMES CONSIDERING TEST ACCURACY BASED ON MY THIN DATA

6.3.1.1 IBD prevalence

IBD prevalence in the tested population within routine practice (THIN data) aligns more closely with that described in the model's sensitivity analysis of a wider population than that in the base case scenario, see Table 6.4. This would suggest reduced cost effectiveness in comparison to the model base case, and that FC

testing is only marginally better than GP assessment alone if all other assumptions were to be true (Table 6.5). However, this is not the case as I will discuss next.

Table 6.4 Main model parameters for base case and sensitivity analysis compared to findings from my THIN analysis in section 4.3

	GP assessment	FC testing (base case)	FC testing wider population (sensitivity analysis)	My THIN study
Population	Symptomatic patients	FC tested patients (those GP would have referred without testing - approx. 25% of symptomatic patients)	FC tested patients (twice as many as GP would have referred – approx. 50% of symptomatic patients)	All FC tested patients
IBD prevalence in population, %	6.3	6.3	3.4	3.5
Proportion referred, %	26.1	11.5	8.9	49.1
Proportion colonoscoped, %	26.1	11.5	8.9	24.7

FC faecal calprotectin, GP general practitioner, IBD inflammatory bowel disease, THIN the Health Improvement Network

6.3.1.2 Test accuracy

The sensitivity of FC testing in the THIN population was similar to the estimate used in the model (Table 6.1). Therefore, the impact of false negative results on the QoL measure and on the costs of treating IBD incorrectly as IBS in clinical practice may be similar to the model findings. Specificity was lower in the THIN population because the clinical question was different. In my THIN study, there were more patients with false positive results who were subsequently referred to colonoscopy. This would result in higher costs according to the model (Table 6.5). However, the true consequence of that in terms of QALY and costs is uncertain because other non-IBD organic conditions might warrant a colonoscopy, which was not considered in the model. Consideration of other inflammatory conditions in the non-IBD group would require costings and QALYs for a considerable number of other conditions and decisions as to how these would be valued. Furthermore, my study showed that GPs correctly refer some patients with a false negative FC test result presumably on other signs and symptoms which was not considered in the model. This illustrates a divergence of clinical practice from theoretical modelling. Therefore, transferring test accuracy measures from secondary care to primary care may be inappropriate if the clinical question is different.

Table 6.5 Possible impact of changing from the original model assumptions to my study outcomes on cost-effectiveness findings

Model parameter	Model assumption (base case)	My analyses	Potential direction of change of costs and consequences of FC testing	Potential direction of change of cost-effectiveness findings (FC testing vs GP assessment)
IBD prevalence, %	6.3	3.5 (Table 4.8)	For a fixed cohort, if only the prevalence is varied: TP, FN ↓ FP, TN ↑ Fewer TP: QALY gains ↓	My analysis is similar to the sensitivity analysis in the wider population (prevalence 3.4%) showing overall reduction in cost savings and reduction in QALY gain: Reduction in cost-effectiveness
Specificity, % (95% CI)	94.0 (73.0 to 99.0)	61.5 (60.2 to 62.7) (Table 4.8)	FP ↑ More patients with positive tests receive unnecessary colonoscopy: overall cost ↑ More adverse events from unnecessary colonoscopy: QALY gains ↓	Reduction in cost-effectiveness
Referral to colonoscopy following FC testing, %	11.5 (test positives only)	21.8 - 51.6 (test positives: 34.6 - 60.8 test negatives: 13.0 - 45.3) (Table 4.8)	More patients receive colonoscopy: Overall cost ↑ Referral of test negative patients (FN patients become TPs and TN patients become FP): FP ↑ FN ↓ Fewer FNs: QALY gain ↑ Many more FPs: QALY gain ↓	Reduction in cost-effectiveness

FN false negative, FP false positive, IBD inflammatory bowel disease, QALY quality adjusted life year

6.3.2 DISCUSSING MODEL OUTCOMES CONSIDERING CLINICIAN BEHAVIOUR

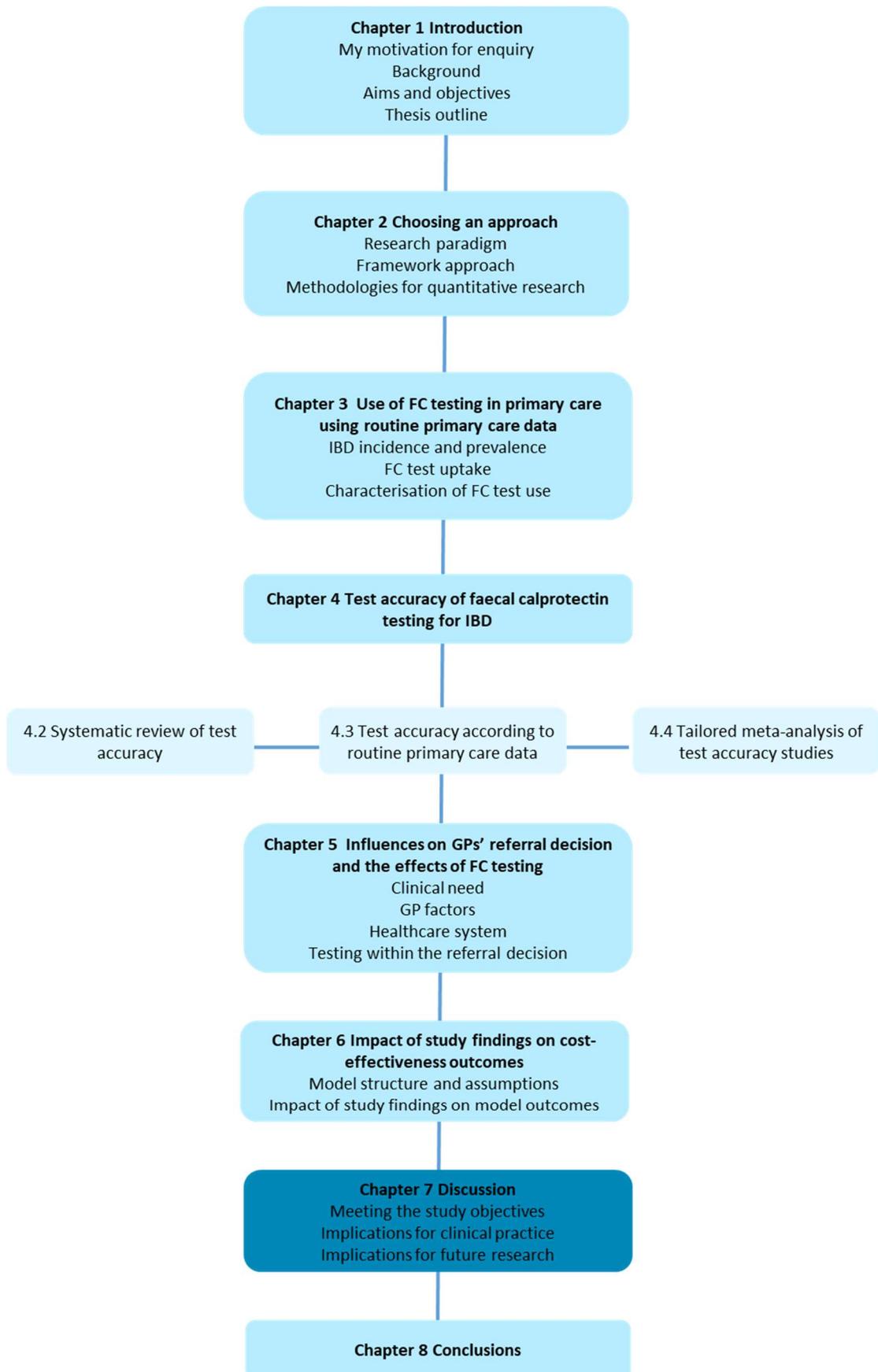
The model did not consider a comparison of GP assessment without FC testing and GP assessment with FC testing but assumed that GPs act purely on the FC test result when FC testing is available. My analysis of the THIN database (Table 4.9) and my qualitative study of GP referral behaviour (see Chapter 5) suggest that this is not the case. Sensitivity and specificity of FC testing may be affected by the referral of patients with negative FC test results due to GPs' uncertainty or GPs' consideration of patient cues in addition to FC levels in referral decisions. Table 6.4 illustrates that the number of tested patients referred to secondary care could be considerably greater than as assumed in the model, where clinical behaviour was assumed to follow FC test results. This difference in referral may not be compensated by consultants' assessments, which could result in a greater proportion of patients receiving colonoscopy than assumed in the model. This would reduce the cost-effectiveness of FC testing (Table 6.5).

Because the costs and consequences of colonoscopy were the driving forces of the model, the cost-effectiveness of FC testing in clinical practice considering only the more traditional patient outcomes is doubtful. My interviews showed that GPs value FC testing for reassurance. We may need to start conversations on how to value test use for reassurance and whether this should be part of cost-effectiveness evaluations.

6.4 Summarising the impact of my study findings on the cost-effectiveness of FC testing

The health economic model showed that FC testing in primary care dominates GP assessment without testing under base case assumptions. The cost effectiveness was driven by the cost of testing, including the FC test and colonoscopy, and possibly the QoL decrement due to adverse events from colonoscopy. The cost savings and QALY gains were reduced in the alternative scenario where more patients with symptoms were tested. This scenario appears to be more relevant to primary care based on prevalence estimates from my THIN data analysis. My findings suggest that the differences in specificity and referral habits would result in a reduction in the cost-effectiveness of FC testing in primary care compared to estimates under model assumptions. Consequently, FC testing may not be cost-effective in primary clinical practice compared to GP assessment alone. While my

study findings were based on a number of assumptions and incomplete data, the evidence on FC test use in primary care suggests that, overall, the current cost-effectiveness estimates could be highly overoptimistic for the primary care setting.



Chapter 7 Discussion

7.1 Chapter overview

In Chapters 3, 4 and 5 I presented and discussed the relevant methods and findings of studies on FC test use (Chapter 3), FC test accuracy (Chapter 4) and FC testing within referral decisions (Chapter 5). In this chapter I reflect on the findings considering my three main objectives. After summarising my contributions to research, I first discuss my findings on test accuracy and FC test use in light of the NICE guidance bringing together insights from my THIN studies and the interview study with GPs. In doing so, I demonstrate how the qualitative data play an essential role in explaining some of the quantitative findings. Then, I consider implications for clinical practice and finish with implications for future research.

7.2 Meeting my study objectives

7.2.1 SUMMARISING MY MAIN AIM AND OBJECTIVES

In my PhD I set out to explore whether a lack of primary care evidence on FC testing has led to assumptions in the cost-effectiveness evaluation that resulted in overoptimistic findings and premature approval of FC testing in primary care. The assumptions related to test use and test accuracy. Thus, I designed studies to address the following three main objectives:

- 1) To investigate the current use of FC testing in primary care and referral of tested patients,
- 2) To estimate the test accuracy of FC testing in primary care and
- 3) To explore influences on GPs' decision to refer in order to explain observed referral patterns following FC testing.

The aim of my research has been to infer from this new evidence which direction the cost-effectiveness outcomes would take by contrasting the evidence with the model inputs and assumptions (Chapter 6).

7.2.2 SUMMARISING MY CONTRIBUTION TO RESEARCH

My research showed that the NICE guidance for FC testing in primary care was based on assumptions that are inaccurate and removed from primary care clinical practice. By taking a secondary care perspective NICE revealed a lack of knowledge and understanding of how primary care functions. The assumptions made by NICE over the potential for FC testing to reduce the costs of colonoscopies have several limitations.

Firstly, the NICE evaluation assumed that test accuracy of FC testing in primary care is consistent with the test accuracy of secondary care studies of selected populations.⁵⁸ I showed in my evaluation of test accuracy that the specificity of FC testing is lower than reported by the NICE evaluation. However, I could not prove that the different disease spectrum in primary care compared to secondary care populations caused biased results. Instead, the overestimation of specificity appeared to be due to the clinical scenario of IBD versus IBS chosen in the NICE evaluation, which is not applicable to clinical practice. The population relevant to primary care is patients with abdominal symptoms rather than patients with a diagnosis of IBD or IBS. Estimates from my test accuracy study of FC testing for IBD versus non-IBD may be more applicable (Chapter 4).

Secondly, the NICE evaluation assumed that FC testing is only undertaken for diagnosis and that GPs would manage patients in line with the FC test results, meaning they would only refer FC test positive patients.⁵⁸ My analyses suggest that these assumptions are incorrect. This included my findings that very few patients followed the anticipated NICE pathway (see Venn diagram in Figure 3.10), that a great proportion of patients with a negative FC test result were referred, that GPs test for reassurance and for legal protection in addition to testing for diagnosis, and that the FC test result is not considered in isolation. My exploration of the role of FC testing in the broader referral decision question (Chapter 5) allowed me to map the outcomes of FC testing alongside the main influences of referral decisions. The results suggested that the referral decision process is too complex to predict the impact that FC testing may have on the referral decision. FC testing cannot override other influences on referral decisions including clinical uncertainty and patient preferences. Furthermore, FC testing may not alleviate all uncertainties during referral decisions and can only provide additional information towards complex referral decisions. Therefore, the assumption that GPs would manage patients in line with the FC test results was too simplistic.

Thirdly, the NICE evaluation assumed that all referred patients would receive colonoscopy and that by reducing the number of referrals, FC testing would have a direct influence on the number of colonoscopies.⁵⁸ This was predicted to result in cost savings. My investigations into the use of FC testing in routine clinical practice showed that only 40% of patients with a positive FC test result received colonoscopy and that 14% of patient with a negative FC test were subsequently colonoscoped (Chapter 3). This suggested that FC testing in primary care does not

directly influence the decisions by specialists over whether to investigate referred patients. Without the expected reduction in colonoscopies I showed that FC testing may not be cost-effective as it is currently used in primary clinical practice (Chapter 6).

As a consequence of the incorrect assumptions in the NICE evaluation, the subsequent NICE guidance on FC testing may not be sufficiently informative and appropriate to promote uptake or consistent use of FC testing. I presented evidence for this demonstrating the variation in the number of FC tests by GP practice and the variation in the number of tests per patient. The data suggested that this may have led to non-compliance with NICE recommendations. This is based on my findings concerning the high number of IBS patients who received FC testing, the low number of eligible patients who received FC testing and the number of patients aged 60 years and over who received FC testing.

In my research I identified some important limitations in the NICE approach to evaluating diagnostic tests for primary care and showed that NICE needs to change its approach in formulating guidance for testing in primary care. I showed that we cannot simply use secondary care evidence to make decisions for primary care.

In the following section I discuss my findings further in the light of the NICE guidance.

7.2.3 TEST ACCURACY OF FC TESTING IN PRIMARY CARE IN LIGHT OF NICE GUIDANCE

7.2.3.1 The effect of patient spectrum on test accuracy

I expected to be able to show that transferring test accuracy measures from secondary care studies to the primary care context will lead to spectrum bias because sensitivity and specificity depend on the spectrum of disease.⁶⁸ While the prevalence of IBD was higher in secondary care studies than primary care studies (Table 7.1), which may suggest a difference in patient spectrum,^{166, 210, 211} my THIN test accuracy estimates were similar to those derived from my meta-analysis of studies from primary and secondary care. However, patient spectrum is only one underlying mechanism that can affect, both, IBD prevalence and test accuracy.⁶⁸ Leeflang et al. (2009) identified other mechanisms which may explain my study findings.⁶⁸ I consider these mechanisms in Table 7.1 for their potential impact on the patient population, and on test accuracy measures in my primary care test accuracy

study compared to secondary care studies. It appears that the spectrum bias, which may lead to lower sensitivity in primary care due to the presence of a bigger pool of milder disease, may have been masked by effects from differential verification. Using follow-up rather than confirmatory testing of test negative patients can underestimate the disease prevalence, overestimate test sensitivity and underestimate test specificity.⁶⁸ This could be, for instance, because patients with IBD are lost during follow-up, or follow-up fails to detect all IBD cases. Furthermore, including tested rather than eligible patients, which led to a higher IBD prevalence in my study (Table 7.1), may have reduced the spectrum effect because the patient spectrum was closer to that found in the secondary care population.

These different mechanisms can change the prevalence and diagnostic accuracy in different directions. This may explain my inability to detect and measure the effect of spectrum bias on the test accuracy estimates. Alternatively, the spectrum effect may not have been large enough to result in noticeable differences in test accuracy estimates. While acknowledging the possibility of bias in my test accuracy estimates, it appears that the NICE approach of using secondary care studies estimated the sensitivity of FC testing in the tested primary care population adequately.

Table 7.1 Impact of mechanisms identified by Leeflang (2009) on the patient population and FC test accuracy in my primary care study population compared to secondary care studies

Mechanism from Leeflang (2009)⁶⁸	Impact on primary care population compared to secondary care population	Impact on sensitivity and specificity in primary care compared to secondary care
<i>Patient spectrum</i> - the range of severity of disease symptoms	<p>Bigger pool of milder disease in my primary care study where FC test may perform sub-optimally, resulting in higher number of false negative results.</p> <p>This difference in patient spectrum is greater in the eligible than tested population. My study included FC tested patients, which may have reduced the spectrum effect on sensitivity.</p>	Sensitivity ↓
<i>Referral filter</i> – prior testing (including assessment) to select patients for further testing	Referral filter is GP assessment. In primary care studies, GP assessment identifies high risk patients for FC testing. In secondary care studies, GP assessment identifies high risk patients for referral. GP's threshold to order FC	See patient spectrum

	<p>testing is likely to be lower than their threshold to refer for colonoscopy. Differences in prevalence in tested and referred populations suggests GPs use different criteria in assessment. The differences result in a different patient spectrum.</p> <table border="1"> <thead> <tr> <th>p_{eligible}</th> <th>p_{tested}</th> <th>p_{result}</th> <th>p_{sec}</th> </tr> </thead> <tbody> <tr> <td>1.4%</td> <td>3.4%</td> <td>3.5%</td> <td>13%</td> </tr> </tbody> </table>	p _{eligible}	p _{tested}	p _{result}	p _{sec}	1.4%	3.4%	3.5%	13%	
p _{eligible}	p _{tested}	p _{result}	p _{sec}							
1.4%	3.4%	3.5%	13%							
<i>Reader expectation</i> – the threshold for declaring a finding as abnormal in response to perceived prevalence	Not applicable as FC testing is not a subjective test	Not applicable								
<i>Biased patient sampling</i> – selection of patients with and without the target condition from different patient populations (case-control studies)	My primary care study, and the secondary care studies included in the model are single gate (cohort) rather than two-gate (case control) designs, which limits this bias. However, there are potentially more patients excluded in my primary care THIN database study because of high proportion of missing FC test results. Tested variables suggest small difference in effect sizes between included and excluded patients, therefore, bias unlikely but there could be effects in unmeasured variables (Table 4.4).	Sensitivity and specificity similar								
<i>Verification bias</i> – verification of disease by different reference tests for test positive and test negative patients	In secondary care studies verification was by colonoscopy. In my primary care study there was differential verification, patients with negative test result received follow-up to confirm disease status and did not all receive colonoscopy. Patients may be lost to follow-up or follow-up may have missed some disease resulting in lower prevalence and lower number of false negatives.	Sensitivity ↑								
<i>Imperfect reference test</i> – verification of disease by an imperfect reference test leading to reference standard misclassification	<p>The reference standard in my primary care study was a coded IBD record rather than colonoscopy. Misclassification of patients was possible due to:</p> <ol style="list-style-type: none"> 1) Missing IBD records (prevalence ↓) leading to some FC positive tests being classified incorrectly as false positives (FP ↑) and some negative tests incorrectly classified as true negatives (TN ↑) 2) Incorrect IBD records (prevalence ↑) leading to some FC negative tests being classified as false negatives (FN ↑) and some positive tests being incorrectly classified as true positive (TP ↑) 	<p>Effect on specificity depends on FC test results</p> <p>Effect on sensitivity depends on FC test results</p>								

FC faecal calprotectin, FN false negative, FP false positive, GP general practitioner, IBD inflammatory bowel disease, IBS irritable bowel syndrome, p prevalence, p_{eligible} prevalence

in eligible patient population, p_{result} prevalence in population with FC test result, p_{sec} mean prevalence of meta-analysed studies from secondary and primary care, p_{tested} prevalence in FC tested population

7.2.3.2 The definition of the non-IBD group

Given the constraints of my THIN test accuracy study discussed in section 4.3.5, I cannot report an undisputable pair of sensitivity and specificity values for FC testing in primary care. However, triangulating findings from the THIN test accuracy study and the tailored meta-analysis showed that the sensitivity of FC testing is high at between 0.93 (0.89 to 0.96) and 0.97 (0.81 to 0.996). In contrast to sensitivity, comparison of the specificity using THIN data, conventional and tailored meta-analysis (section 4.4.4.2) showed greater variation and was lower than the NICE estimate (Table 6.7). I believe that this was due to the various definitions of the non-IBD group across studies. Studies included in the meta-analyses reported a range of conditions other than IBS that were classified as non-IBD. Collectively, they included nonsteroidal anti-inflammatory drug enteropathy, adenoma, infectious colitis, CRC, polyps, diverticulitis, appendicitis, diversion proctitis, bacterial overgrowth, collagenous colitis, small bowel obstruction, bleeding ulcer, gastritis, haemorrhoids, ischaemic colitis, reflux esophagitis, microscopic colitis, lactose intolerance, coeliac disease and arthritis.^{63, 131, 137, 138, 140} The proportion of alternative conditions within the non-IBD group ranged from 4% (16/399)⁶³ to 61% (422/694).¹³⁷ I was, therefore, unable to predict the proportion of alternative diagnoses in non-IBD patients in my THIN test accuracy study and the meaning of these false positives in the clinical context may be contentious. The range of reported alternative conditions also shows that an assessment of the test assuming an IBD/IBS population, which underlies the NICE assessment, is clearly not appropriate for routine primary care. The range of possible alternative diagnoses also demonstrates the difficulty of defining the non-IBD group for a retrospective study.

Following my studies on the test accuracy of FC testing in primary care, some uncertainty remains concerning the test's specificity. Further research to evaluate the clinical and cost implications of the detection of the other conditions listed above may be appropriate.

7.2.3.3 The FC pathway in primary care

The eligible patient population for testing, the tested population and the referred population may not neatly follow the expected testing pathway in the NICE guidance. I discovered two examples that highlight this. Firstly, my THIN test accuracy study showed that an unexpectedly high proportion of IBS patients were tested (section 4.3.4.2). My interview study suggests that this may be the result of testing for reassurance, which is not listed as a reason for testing in the test pathway. This may lead to additional false positive and true negative results. Consequently, FC testing could have a lower positive predictive value and a higher negative predictive value in primary care than predicted by NICE guidance. Secondly, the THIN data revealed that GPs did not refer on test results alone (Table 4.9). In the interviews GPs explained that an FC test is not used in isolation from patient assessment and that GPs refer for reasons other than diagnosis. A negative result was often not sufficient to prevent referral and GPs used other cues from patient assessment in the decision about whether to refer. Referral of patients with a negative FC test result reduces the overall specificity of the FC testing pathway in primary care as the majority of FC negative referrals would be IBD negative. Referral of false negatives could result in higher sensitivity of FC testing in combination with the GPs' assessment than FC testing alone. However, my findings appear to suggest that GP assessment also delays timely referral for some IBD patients following a positive test. Consequently, the expected test pathway based on the secondary care perspective has not translated well into primary care clinical practice as it ignores alternative motivations for testing and referral rooted in the different nature and practice of primary care. Clearly, this has an effect on test accuracy of FC testing in primary care. FC testing in combination with GP assessment is a complex intervention, which should be evaluated together to inform NICE guidance. Whilst cost-effectiveness models are not intended to exactly replicate reality, in this case I consider a model more closely approximating primary care practice would be appropriate.

7.2.3.4. Lessons learned from my studies on test accuracy of FC testing in primary care

Overall, my analysis of routine data did not provide a clear answer on the test accuracy of FC testing in the primary care population. However, it provided an answer to the question of how the FC test performs in routine clinical practice considering FC testing alongside GP assessment. Taking the best estimate of my primary analysis, FC testing, when used by GPs, did not miss many cases of IBD

(NPV 0.996, 95% CI 0.993 to 0.997). GPs could, therefore, use the test confidently in ruling out IBD if the test is negative. This is important because my GP interviews showed that the test is often used for reassurance by GPs. However, the positive predictive value of FC testing is poor because FC is a marker of intestinal inflammation rather than a specific marker for IBD. Consequently, the FC test does not perform as well in identifying IBD on a background of non-IBD in contrast to NICE guidance which considered a background of IBS. The data also suggest that the higher threshold of 100µg/g may be appropriate for routine primary care, considering that the test is not used in isolation and that the objective is to reduce the number of unnecessary colonoscopies. This is under the assumption that missed IBD patients would re-present and be re-considered for referral within an appropriate time frame. However, the latest FC pathway recommends two cut-offs (100µg/g and 250µg/g) with an intermediate test range for retesting.¹¹⁸ The additional IBD cases missed at initial testing with this strategy may be inappropriately high (65/210, 31%). I have not considered the feasibility of this latest pathway recommendation with a retesting strategy in primary care in this thesis but it is likely to be more difficult to interpret for general practitioners.

7.2.4 FC TEST USE IN PRIMARY CARE IN LIGHT OF NICE GUIDANCE

7.2.4.1 The test population in primary care

In an ideal test accuracy study all patients with eligible symptoms are tested.¹⁴⁹ My test accuracy study included FC tested patients. The Venn diagram (Figure 3.10) showed that there is little overlap between eligible patients according to NICE guidance and FC tested patients. As a consequence, I am unable to define the tested population sufficiently to describe who GPs test using the THIN data. The NICE guidance states that FC testing aids the differential diagnosis of IBD and IBS in patients with ongoing abdominal symptoms when referral is considered.⁵⁹ This is subjective. My interviews with GPs revealed that this results in variation between GPs in who to test and why. GPs use FC testing for diagnostic uncertainty, reassurance and medico-legal protection. However, guidelines only stipulate diagnostic uncertainty as a reason. Without clear guidance on who to test and how to act on test results, testing can be variable and potentially inappropriate. I see one challenge in the fact that guidelines and cost-effectiveness evaluations have a public health perspective while GPs need to translate these to unique patient encounters requiring individual decisions. A second challenge is that guidelines describe the test pathway for IBD rather than for abdominal pain. While it is straight forward to identify a pathway for IBD starting from the endpoint, it is more

complicated from the GP's perspective from the beginning since they start with abdominal pain and other symptoms which could result in a number of different diagnoses and outcomes. A third challenge is that guidelines are not sufficiently prescriptive to support a non-referral. Fear of litigation and patient preference appear to be common reasons for testing and referral. These are not part of the test pathway and lead to a different testing population than predicted in NICE guidance, which was built on cost-effectiveness assumptions taking a secondary care perspective. NICE appear to have a lack of understanding of how general practice functions. The NICE Diagnostic Assessment Programme should adopt a primary care perspective in the assessment of tests for primary care.

7.2.4.2 Impact of FC testing on management decisions

NICE guidance approved FC testing to reduce the number of unnecessary referrals to colonoscopy.⁵⁹ The model underlying the NICE guidance assumed that GPs would only refer test positive patients. My THIN data revealed that nearly 50% of FC test negative patients were referred and this is in line with previously published evidence from primary care (section 3.5.4.2). Interviews with GPs clearly showed that an FC test does not influence a referral decision directly and that a negative FC test is often not sufficient evidence to prevent a referral. I, consequently, questioned whether FC testing could have the desired impact on the number of patients referred and colonoscoped as predicted in the NICE guidance. However, lacking data on referral without FC testing, I was not able to study the true impact that FC testing may have. One primary care study recorded the GPs' rhetoric referral decision by asking them whether they would have referred without FC testing before an FC test was ordered.⁶⁴ They compared this expected referral with the observed referral following FC testing. They reported that 35% (279/789 tested patients) of referrals to colonoscopy were saved.⁶⁴ Whether this can be achieved under non-study conditions remains uncertain. Audits of FC test pathways report contradicting findings in terms of colonoscopy following the introduction of FC testing.^{119, 212} GPs in my interviews indicated that referrals are often made when uncertainty remains, for difficult patients and for advice and support. Clearer information on the accuracy of FC testing and pathways on how to act on FC test results in NICE guidance may help in alleviating some of the uncertainty resulting in referrals.

7.2.5 GENERALISABILITY OF STUDY FINDINGS

My THIN studies on test accuracy and FC test use were based on routine data collected until January 2017. My study on FC test uptake revealed that by 2016,

four years after the publication of the NICE FC guidance, only 67% of GP practices had started to use FC testing (section 3.5.2). My GP interviews showed that FC testing was still not available to all GP practices by 2018. And some GPs had not used the test even though it was available to them. The theory of diffusion of innovation suggests that adoption is supported by information on the technology's consequences, advantages and disadvantages.¹¹⁷ Furthermore, the communication channel is of importance.¹¹⁷ Some GPs admitted in the interviews that they had not heard of FC testing. The publication of NICE guidance may not be sufficient to inform the majority of GPs as the majority will not look to scientific studies but to subjective evaluations and the experiences of near-peers.¹¹⁷ In the GP interviews the endorsement of tests by local consultants and opinion leaders was clearly described as one of the requirements for testing. This needs to be considered in efforts to promote the use of FC testing in primary care and to influence the time it takes for FC testing to diffuse fully into clinical practice. According to the diffusion of innovation theory, early adopters are different to late adopters in terms of their ability to tolerate uncertainty and their access to resources.¹¹⁷ Therefore, my study findings may look different if the studies were repeated once FC testing is fully established in primary care, GPs routinely use the test and have had opportunity to gain confidence and experience with testing and with their management decisions following testing.

7.3 Implications for clinical practice

My findings on test accuracy suggest that, because of the low prevalence of IBD, FC testing has a high negative predictive value in primary care and is reliable in excluding IBD. However, GPs need to be aware of its low positive predictive value and interpret positive results with caution as the IBD pathway may not be the most appropriate one for their patients.

My findings on the FC test use suggest that there is a discrepancy between the motivation for approval of FC testing from the public health perspective and the perceived advantage of FC testing by individual GPs. The intended benefit of FC testing was to reduce unnecessary colonoscopies in patients where IBD can be ruled out with the help of FC testing. However, some GPs see FC testing as a tool to provide reassurance to IBS patients and use it as a screening tool in abdominal symptoms more widely. Failing to provide that reassurance, which could be due to GPs' lack of knowledge of or confidence in FC test accuracy, results in high

referrals of test negatives. Consequently, the cost-effectiveness result that was based on the public health perspective is not applicable to FC testing as it is currently used in primary clinical practice. FC testing may not be cost-effective in primary care as it is currently implemented.

My findings on the requirements for FC testing suggest that GPs generally welcome tests in primary care. However, information provided with a new test is not standardised and mostly insufficient resulting in variable and inconsistent use. Clearer guidance on who to test and how to act on test results, alongside information on consequences of testing is needed. Furthermore, details about the positive and negative predictive values in primary care and the limitations of FC testing may prevent inappropriate testing, reduce variability in testing and improve confidence in test use and interpretation. Finally, more efficient ways to disseminate the information should be utilised to enable adoption more widely.

7.4 Implications for future research

My PhD research contributes new knowledge towards the test evaluation of faecal calprotectin testing for primary care. It also provides novel insights into methods for test evaluation using routine electronic health records and triangulating test accuracy estimates using tailored meta-analysis and routine data. I have shown that conventional patient outcomes including diagnosis and treatment response are not adequate for test evaluations in primary care because they fail to cover the range of reasons for testing and referral and their consequences. My research has also opened up new research questions and highlighted shortcomings in our current knowledge base. These include:

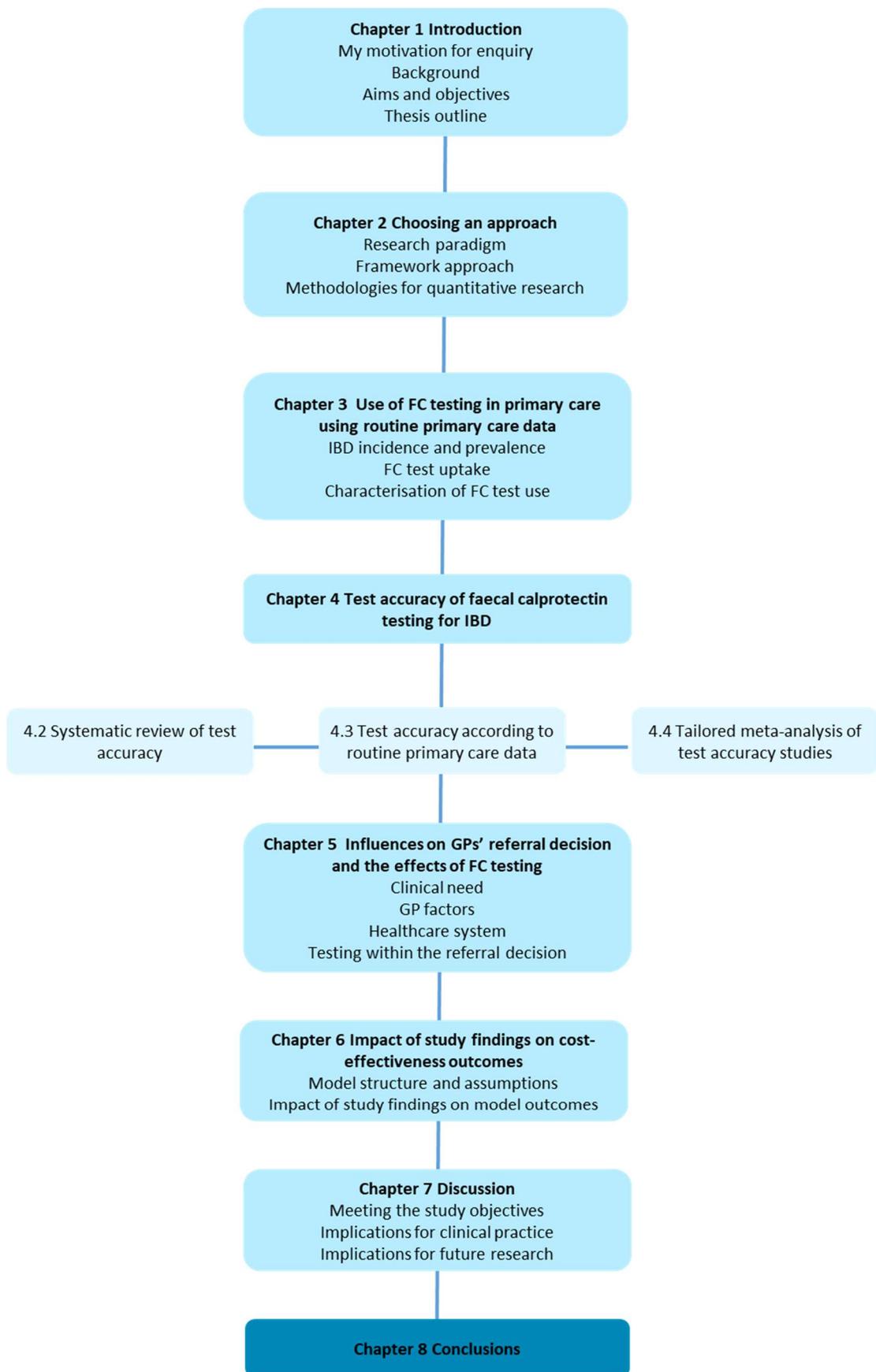
Firstly, there is a need to update the current FC cost-effectiveness model for primary care once test use is established in primary care and referral rates have been studied. The model needs to consider the clinical question relevant to primary care and requires a decision on how to deal with non-IBD positive tests. It also needs to be able to compare accurate numbers of referral to secondary care with and without FC testing.

Secondly, we need to understand to what extent treatment is delayed for false negatives and what the true proportion of false negatives is when the new threshold

is implemented. Related to that is the need to understand how feasible a retesting strategy is using two thresholds in primary care clinical practice.

Thirdly, an understanding of the differences in patients with and without a recorded FC test result is needed to estimate the bias in studies of routine electronic health care records of test results. This could extend research, including my PhD research, into how to measure, quantify and adjust for bias in studies using routine electronic health records adding to the validity and reputation of studies using routine data.

Finally, health policy research is needed into which outcomes should be considered in test evaluations in primary care and how they should be captured and valued in cost-effectiveness analyses ensuring that a primary care perspective is taken.



Chapter 8 Conclusions

The aim of my thesis has been to establish new primary care evidence for the test accuracy and use of FC testing and to explore how this varies with assumptions in the cost-effectiveness model which led to the national approval of FC testing in primary care.

In Chapter 1 I provided an introduction and some clinical background on the conditions and the test of interest. In detailing the lack of primary care evidence in the evaluation of FC testing I laid out the rationale for this thesis.

In Chapter 2 I presented the justification for the main methodologies: the systematic review and meta-analysis to summarise and synthesis the current evidence, the retrospective cohort study of routine electronic health data to examine real world evidence on FC test use and test accuracy and the interview study to explore GPs views on referral and testing. I discussed these methodologies with respect to the epistemological and ontological paradigms that I believe underlie the research questions.

In Chapter 3 I investigated the FC test use in primary clinical practice. I showed that uptake has been slow and that FC test use is inconsistent. Patients following the anticipated NICE pathway (symptoms followed by FC test followed by IBD diagnosis) were in a minority and the considerable referral of FC negative patients only resulted in a small number of IBD diagnoses.

In Chapter 4 I focused on the test accuracy of FC testing to distinguish IBD from non-IBD using three different methods. A conventional meta-analysis of published studies, a tailored meta-analysis of studies deemed applicable to the primary care setting and a test accuracy study of routine primary care data resulted in similarly high sensitivity of FC testing. The evidence on specificity was less certain and lower than that for the clinical question of IBD versus IBS. This was because of the considerable number of false positives which, depending on the definition of non-IBD, could include a range of inflammatory conditions.

In Chapter 5 I provided an explanation to the questions 1) why FC negative patients in routine practice are referred, 2) who GPs order an FC test for and 3) for what

reasons. The interviews with GPs revealed that factors other than FC test results are more influential on the individual referral decision. GPs test and refer for reassurance which is not always met by FC testing. Therefore, a negative FC test result is often not sufficient in preventing referral in cases of uncertainty.

In Chapter 6 I compared my findings with the assumptions of the cost-effectiveness model and built the case that the cost-effectiveness outcomes were overoptimistic because the test is not used as anticipated in routine primary care.

In Chapter 7 I discussed my study findings in the context of the main objectives of my thesis. The two main limitations, which caution my interpretation of the test accuracy estimates, were the proportion of excluded patients due to missing variables and my inability to confirm non-IBD cases using the information available to me in the routine data.

Considering the study findings and the limitations discussed throughout my dissertation I conclude:

- 1) THIN data are useful for evaluations of tests established in routine general practice to answer pragmatic test accuracy questions. The data can provide evidence on test use and on the impact of testing on clinical decision making. However, in the interpretation of findings researchers need to recognise that test accuracy studies using routine data evaluate a combination of test accuracy and GP assessment. Furthermore, limitations and assumptions of the study and their likely impact on study results need to be discussed. Triangulation using other methods can be used to place more confidence in test accuracy outcomes.
- 2) Validation of the method of tailored meta-analysis with routine data showed that the method has got some limitations that could be addressed by adjusting the method to include all available studies weighted by their probability to be applicable to the study setting and by incorporating the constraints that defined the applicable region into the statistical model that produces the summary estimates. This would ensure that tailored summary estimates are always in the applicable region.
- 3) I could not prove that transferring test accuracy measures from the secondary to the primary care setting results in spectrum bias. That does not mean it is safe to use secondary care data for decisions in primary care. FC

testing has a higher specificity in the constructed population of IBD and IBS patients than in a wider population. This means transferring test accuracy measures between settings needs to consider the applicability of the population under study. A study population from explanatory test accuracy studies is unlikely to be relevant to clinical practice.

- 4) The current cost-effectiveness model is not applicable to current practice for two main reasons. Firstly, the non-IBD group is fundamentally different from the IBS group used in the model. Without extensive knowledge of the proportions and types of conditions other than IBS that are important in the primary care testing decision, an adjustment of the costs and consequences of the model is not feasible. Secondly, my findings suggest that clinical behaviour cannot be simply predicted based on test results. FC testing in primary care requires a new cost-effectiveness evaluation. The NICE Diagnostic Assessment Programme should adopt a primary care perspective in the assessment of tests for primary care which needs to include an understanding of how general practice functions.
- 5) In order to achieve the predicted benefits in cost savings and patient outcomes, either cost-effectiveness evaluations need to consider more pragmatic evidence on test use, or guidelines based on cost-effectiveness models need to stipulate more clearly when and how to use the test.

Overall, in my opinion, FC testing is a useful test for primary care and will play an important role in the assessment and management of patients with abdominal symptoms. However, its current use in primary care has not been evaluated in formal cost-effectiveness analyses. FC testing in primary care is currently unlikely to be cost-effective. Test evaluations for primary care need to consider the GPs' decision-making processes for testing and referral decisions. When developing national guidance models should more closely represent general practice, rather than simply extrapolating from secondary care.

References

1. Wallander MA, Johansson S, Ruigomez A, Garcia Rodriguez LA. Unspecified abdominal pain in primary care: the role of gastrointestinal morbidity. *Int J Clin Pract*. 2007;61(10):1663-70. doi: 10.1111/j.1742-1241.2007.01529.x.
2. Thompson WG, Heaton KW, Smyth GT, Smyth C. Irritable bowel syndrome in general practice: prevalence, characteristics, and referral. *Gut*. 2000;46(1):78-82. doi: 10.1136/gut.46.1.78.
3. Viniol A, Keunecke C, Biroga T, Stadje R, Dornieden K, Bosner S, Donner-Banzhoff N, Haasenritter J, Becker A. Studies of the symptom abdominal pain: a systematic review and meta-analysis. *Fam Pract*. 2014;31(5):517-29. doi: 10.1093/fampra/cmu036.
4. Hungin AP, Whorwell PJ, Tack J, Mearin F. The prevalence, patterns and impact of irritable bowel syndrome: an international survey of 40,000 subjects. *Aliment Pharmacol Ther*. 2003;17(5):643-50. doi: 10.1046/j.1365-2036.2003.01456.x.
5. Oberndorff-Klein Woolthuis AH, Brummer RJ, de Wit NJ, Muris JW, Stockbrugger RW. Irritable bowel syndrome in general practice: an overview. *Scand J Gastroenterol Suppl*. 2004(241):17-22. doi: 10.1080/00855920410010942.
6. NHS Technology Adoption Centre. Faecal calprotectin: faecal calprotectin testing in primary care. 2013 17/12/2015 [cited 17/12/2015]; Available from: <http://www.calprotectin.co.uk/case-studies/faecal-calprotectin-testing-in-primary-care/>.
7. El-Serag HB, Pilgrim P, Schoenfeld P. Systemic review: natural history of irritable bowel syndrome. *Aliment Pharmacol Ther*. 2004;19(8):861-70. doi: 10.1111/j.1365-2036.2004.01929.x.
8. Chey WD, Kurlander J, Eswaran S. Irritable bowel syndrome: a clinical review. *JAMA*. 2015;313(9):949-58. doi: <https://dx.doi.org/10.1001/jama.2015.0954>.
9. Hahn B, Watson M, Yan S, Gunput D, Heuwerker J. Irritable bowel syndrome symptom patterns: frequency, duration, and severity. *Dig Dis Sci*. 1998;43(12):2715-8. doi: <https://doi.org/10.1023/A:1026663613695>.
10. Waller SL, Misiewicz JJ. Prognosis in the irritable-bowel syndrome: a prospective study. *Lancet*. 1969;2(7624):754-6. doi: [https://doi.org/10.1016/S0140-6736\(69\)90475-9](https://doi.org/10.1016/S0140-6736(69)90475-9).
11. Nellesen D, Chawla A, Oh DL, Weissman T, Lavins BJ, Murray CW. Comorbidities in patients with irritable bowel syndrome with constipation or chronic idiopathic constipation: a review of the literature from the past decade. *Postgrad Med*. 2013;125(2):40-50. doi: 10.3810/pgm.2013.03.2640.
12. Canavan C, West J, Card T. The epidemiology of irritable bowel syndrome. *Clin Epidemiol*. 2014;6:71-80. doi: 10.2147/CLEP.S40245.
13. Manning AP, Thompson WG, Heaton KW, Morris AF. Towards positive diagnosis of the irritable bowel. *BMJ*. 1978;2(6138):653-4. doi: 10.1136/bmj.2.6138.653.
14. Kruis W, Thieme C, Weinzierl M, Schussler P, Holl J, Paulus W. A diagnostic score for the irritable bowel syndrome: its value in the exclusion of organic disease. *Gastroenterology*. 1984;87(1):1-7. doi: [https://doi.org/10.1016/0016-5085\(84\)90119-7](https://doi.org/10.1016/0016-5085(84)90119-7).
15. Lacy BE, Patel NK. Rome criteria and a diagnostic approach to irritable bowel syndrome. *J Clin Med*. 2017;6(11):1-8. doi: 10.3390/jcm6110099.
16. Ford AC, Bercik P, Morgan DG, Bolino C, Pintos-Sanchez MI, Moayyedi P. Validation of the Rome III criteria for the diagnosis of irritable bowel syndrome in secondary care. *Gastroenterology*. 2013;145(6):1262-70. doi: 10.1053/j.gastro.2013.08.048.
17. Thompson WG, Heaton KW, Smyth GT, Smyth C. Irritable bowel syndrome: the view from general practice. *Eur J Gastroenterol Hepatol*. 1997;9(7):689-92. doi: 10.1097/00042737-199707000-00008.
18. National Institute of Health and Care Excellence. Irritable bowel syndrome in adults: diagnosis and management [CG61]. 2017 [cited 13/08/2018]; Available from: <https://www.nice.org.uk/guidance/cg61/chapter/1-Recommendations#diagnosis-of-ibs>.
19. Ungaro R, Mehandru S, Allen PB, Peyrin-Biroulet L, Colombel JF. Ulcerative colitis. *Lancet*. 2017;389(10080):1756-70. doi: 10.1016/s0140-6736(16)32126-2.
20. Torres J, Mehandru S, Colombel JF, Peyrin-Biroulet L. Crohn's disease. *Lancet*. 2017;389(10080):1741-55. doi: 10.1016/s0140-6736(16)31711-1.
21. Chan D, Robbins H, Rogers S, Clark S, Poullis A. Inflammatory bowel disease and exercise: results of a Crohn's and Colitis UK survey. *Frontline Gastroenterol*. 2014;5(1):44-8. doi: 10.1136/flgastro-2013-100339.

22. Probert CS, Jayanthi V, Hughes AO, Thompson JR, Wicks AC, Mayberry JF. Prevalence and family risk of ulcerative colitis and Crohn's disease: an epidemiological study among Europeans and south Asians in Leicestershire. *Gut*. 1993;34(11):1547-51. doi: 10.1136/gut.34.11.1547.
23. Rubin GP, Hungin AP, Kelly PJ, Ling J. Inflammatory bowel disease: epidemiology and management in an English general practice population. *Aliment Pharmacol Ther*. 2000;14(12):1553-9. doi: 10.1046/j.1365-2036.2000.00886.x.
24. Lewis JD, Brensinger C, Bilker WB, Strom BL. Validity and completeness of the General Practice Research Database for studies of inflammatory bowel disease. *Pharmacoepidemiol Drug Saf*. 2002;11(3):211-8. doi: 10.1002/pds.698.
25. Stone MA, Mayberry JF, Baker R. Prevalence and management of inflammatory bowel disease: a cross-sectional study from central England. *Eur J Gastroenterol Hepatol*. 2003;15(12):1275-80. doi: 10.1097/01.meg.0000085500.01212.e2.
26. Frolkis AD, Dykeman J, Negron ME, Debruyjn J, Jette N, Fiest KM, Frolkis T, Barkema HW, Rioux KP, Panaccione R, Ghosh S, Wiebe S, Kaplan GG. Risk of surgery for inflammatory bowel diseases has decreased over time: a systematic review and meta-analysis of population-based studies. *Gastroenterology*. 2013;145(5):996-1006. doi: 10.1053/j.gastro.2013.07.041.
27. Weimers P, Munkholm P. The natural history of IBD: lessons learned. *Curr Treat Options Gastroenterol*. 2018;16(1):101-11. doi: 10.1007/s11938-018-0173-3.
28. Bernstein CN, Wajda A, Blanchard JF. The clustering of other chronic inflammatory diseases in inflammatory bowel disease: a population-based study. *Gastroenterology*. 2005;129(3):827-36. doi: 10.1053/j.gastro.2005.06.021.
29. Molnar T, Annahazi A. Pathogenesis of ulcerative colitis and Crohn's disease: similarities, differences and a lot of things we do not know yet. *J Clin Cell Immunol*. 2014;5:253. doi: 10.4172/2155-9899.1000253.
30. Vavricka SR, Spigaglia SM, Rogler G, Pittet V, Michetti P, Felley C, Mottet C, Braegger CP, Rogler D, Straumann A, Bauerfeind P, Fried M, Schoepfer AM. Systematic evaluation of risk factors for diagnostic delay in inflammatory bowel disease. *Inflamm Bowel Dis*. 2012;18(3):496-505. doi: 10.1002/ibd.21719.
31. Stapley SA, Rubin GP, Alsina D, Shephard EA, Rutter MD, Hamilton WT. Clinical features of bowel disease in patients aged <50 years in primary care: a large case-control study. *Br J Gen Pract*. 2017;67(658):e336-e44. doi: 10.3399/bjgp17X690425.
32. Satsangi J, Silverberg MS, Vermeire S, Colombel JF. The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications. *Gut*. 2006;55(6):749-53. doi: 10.1136/gut.2005.082909.
33. Li XG, Lu YM, Gu F, Yang XL. [Fecal calprotectin in differential diagnosis of irritable bowel syndrome]. *Beijing Da Xue Xue Bao*. 2006;38(3):310-3.
34. Carter M, Lobo A, Travis S. Guidelines for the management of inflammatory bowel disease in adults. *Gut*. 2004;53(Suppl 5):v1-v16. doi: 10.1136/gut.2004.043372.
35. Alexakis C, Davies G, Stephens J, Clark S, Rogers S, Poullis A. Perspectives and attitudes of young patients with inflammatory bowel disease: symptoms, burden of disease and communication with their healthcare professionals. *Frontline Gastroenterol*. 2014;5(3):197-202. doi: 10.1136/flgastro-2013-100400.
36. Porter CK, Cash BD, Pimentel M, Akinseye A, Riddle MS. Risk of inflammatory bowel disease following a diagnosis of irritable bowel syndrome. *BMC Gastroenterol*. 2012;12:55. doi: 10.1186/1471-230x-12-55.
37. Gracie DJ, Ford AC. Irritable bowel syndrome-type symptoms are associated with psychological comorbidity, reduced quality of life, and health care use in patients with inflammatory bowel disease. *Gastroenterology*. 2017;153(1):324-5. doi: 10.1053/j.gastro.2017.05.037.
38. van den Heuvel-Janssen HAM, Borghouts JAJ, Muris JWM, Koes BW, Bouter LM, Knottnerus JA. Chronic non-specific abdominal complaints in general practice: a prospective study on management, patient health status and course of complaints. *BMC Fam Pract*. 2006;7:12. doi: 10.1186/1471-2296-7-12.
39. Janssen HA, Borghouts JA, Muris JW, Metsemakers JF, Koes BW, Knottnerus JA. Health status and management of chronic non-specific abdominal complaints in general practice. *Br J Gen Pract*. 2000;50(454):375-9.
40. Kok L, Elias SG, Witterman BJ, Goedhard JG, Muris JW, Moons KG, de Wit NJ. Diagnostic accuracy of point-of-care fecal calprotectin and immunochemical occult blood

- tests for diagnosis of organic bowel disease in primary care: the cost-effectiveness of a decision rule for abdominal complaints in primary care (CEDAR) study. *Clin Chem*. 2012;58(6):989-98. doi: <http://dx.doi.org/10.1373/clinchem.2011.177980>.
41. Eskeland SL, Dalen E, Sponheim J, Lind E, Brunborg C, de Lange T. European panel on the appropriateness of gastrointestinal endoscopy II guidelines help in selecting and prioritizing patients referred to colonoscopy: a quality control study. *Scand J Gastroenterol*. 2014;49(4):492-500. doi: 10.3109/00365521.2014.886715.
 42. Tavabie OD, Hughes SA, Loganayagam A. The role of faecal calprotectin in the differentiation of organic from functional bowel disorders. *Br J Gen Pract*. 2014;64(628):595-6. doi: <https://dx.doi.org/10.3399/bjgp14X682525>.
 43. van Rheenen PF. Faecal calprotectin for screening of patients with suspected inflammatory bowel disease: diagnostic meta-analysis. *BMJ*. 2010;341:c3369. doi: 10.1136/bmj.c3369.
 44. Dabritz J, Musci J, Foell D. Diagnostic utility of faecal biomarkers in patients with irritable bowel syndrome. *World J Gastroenterol*. 2014;20(2):363-75. doi: 10.3748/wjg.v20.i2.363.
 45. Schroder O, Naumann M, Shastri Y, Povse N, Stein J. Prospective evaluation of faecal neutrophil-derived proteins in identifying intestinal inflammation: combination of parameters does not improve diagnostic accuracy of calprotectin. *Aliment Pharmacol Ther*. 2007;26(7):1035-42. doi: 10.1111/j.1365-2036.2007.03457.x.
 46. Roseth AG, Schmidt PN, Fagerhol MK. Correlation between faecal excretion of indium-111-labelled granulocytes and calprotectin, a granulocyte marker protein, in patients with inflammatory bowel disease. *Scand J Gastroenterol*. 1999;34(1):50-4. doi: 10.1080/00365529950172835.
 47. Roseth AG, Fagerhol MK, Aadland E, Schjonsby H. Assessment of the neutrophil dominating protein calprotectin in feces: a methodologic study. *Scand J Gastroenterol*. 1992;27(9):793-8. doi: 10.3109/00365529209011186.
 48. Roseth AG, Aadland E, Jahnsen J, Raknerud N. Assessment of disease activity in ulcerative colitis by faecal calprotectin, a novel granulocyte marker protein. *Digestion*. 1997;58(2):176-80. doi: 10.1159/000201441.
 49. Tibble J, Teahon K, Thjodleifsson B, Roseth A, Sigthorsson G, Bridger S, Foster R, Sherwood R, Fagerhol M, Bjarnason I. A simple method for assessing intestinal inflammation in Crohn's disease. *Gut*. 2000;47(4):506-13. doi: 10.1136/gut.47.4.506.
 50. Kennedy NA, Clark A, Walkden A, Chang JC, Fasci-Spurio F, Muscat M, Gordon BW, Kingstone K, Satsangi J, Arnott ID, Lees CW. Clinical utility and diagnostic accuracy of faecal calprotectin for IBD at first presentation to gastroenterology services in adults aged 16-50 years. *J Crohns Colitis*. 2015;9(1):41-9. doi: <https://dx.doi.org/10.1016/j.crohns.2014.07.005>.
 51. Pavlidis P, Chedgy FJ, Tibble JA. Diagnostic accuracy and clinical application of faecal calprotectin in adult patients presenting with gastrointestinal symptoms in primary care. *Scand J Gastroenterol*. 2013;48(9):1048-54. doi: <http://dx.doi.org/10.3109/00365521.2013.816771>.
 52. Jang HW, Kim HS, Park SJ, Hong SP, Kim TI, Kim WH, Cheon JH. Accuracy of three different fecal calprotectin tests in the diagnosis of inflammatory bowel disease. *Intest Res*. 2016;14(4):305-13. doi: <https://dx.doi.org/10.5217/ir.2016.14.4.305>.
 53. Konikoff MR, Denson LA. Role of fecal calprotectin as a biomarker of intestinal inflammation in inflammatory bowel disease. *Inflamm Bowel Dis*. 2006;12(6):524-34. doi: 10.1097/00054725-200606000-00013.
 54. Banerjee A, Srinivas M, Eyre R, Ellis R, Waugh N, Bardhan KD, Basumani P. Faecal calprotectin for differentiating between irritable bowel syndrome and inflammatory bowel disease: a useful screen in daily gastroenterology practice. *Frontline Gastroenterol*. 2015;6(1):20-6. doi: <http://dx.doi.org/10.1136/flgastro-2013-100429>.
 55. Turvill J, O'Connell S, Brooks A, Bradley-Wood K, Laing J, Thiagarajan S, Hammond D, Turnock D, Jones A, Sood R, Ford A. Evaluation of a faecal calprotectin care pathway for use in primary care. *Prim Health Care Res Dev*. 2016;17(5):428-36. doi: <https://dx.doi.org/10.1017/S1463423616000049>.
 56. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Med Decis Making*. 1991;11(2):88-94. doi: 10.1177/0272989x9101100203.

57. Ferrante di Ruffano L, Hyde CJ, McCaffery KJ, Bossuyt PM, Deeks JJ. Assessing the value of diagnostic tests: a framework for designing and evaluating trials. *BMJ*. 2012;344:e686. doi: 10.1136/bmj.e686.
58. Waugh N, Cummins E, Royle P, Kandala NB, Shyangdan D, Arasaradnam R, Clar C, Johnston R. Faecal calprotectin testing for differentiating amongst inflammatory and non-inflammatory bowel diseases: systematic review and economic evaluation. *Health Technol Assess*. 2013;17(55):xv-xix, 1-211. doi: <http://dx.doi.org/10.3310/hta17550>.
59. National Institute for Health and Care Excellence. Faecal calprotectin diagnostic tests for inflammatory diseases of the bowel: NICE diagnostics guidance [DG11]. 2013 [cited 17/12/2015]; Available from: <https://www.nice.org.uk/guidance/dg11>.
60. McFarlane M, Chambers S, Malik A, Lee B, Sung E, Nwokolo C, Waugh N, Arasaradnam R. Is NICE too optimistic about savings from normal faecal calprotectin results? *J Gastroenterol Hepatol Res*. 2016;5(1):1895-8. doi: 10.17554/j.issn.2224-3992.2016.05.580
61. Boyd A, Peplow C, Dhaliwal A, O'Rourke J, Davies J, Milestone A. Primary care faecal calprotectin pathway: a three year analysis. *Colorectal Disease*. 2016;18:27. doi: <http://dx.doi.org/10.1111/codi.13444>.
62. Hogberg C, Karling P, Rutegard J, Lilja M. Diagnosing colorectal cancer and inflammatory bowel disease in primary care: the usefulness of tests for faecal haemoglobin, faecal calprotectin, anaemia and iron deficiency. A prospective study. *Scand J Gastroenterol*. 2017;52(1):69-75. doi: <https://dx.doi.org/10.1080/00365521.2016.1228120>.
63. Conroy S, Hale MF, Cross SS, Swallow K, Sidhu RH, Sargur R, Lobo AJ. Unrestricted faecal calprotectin testing performs poorly in the diagnosis of inflammatory bowel disease in patients in primary care. *J Clin Pathol*. 2018;71:316-22. doi: <http://dx.doi.org/10.1136/jclinpath-2017-204506>.
64. Walker G, Moore L, Heerasing N, Hendy P, Perry M, McDonald T, Debenham T, Bethune R, Bewshea C, Hyde C, Heap G, Singh A, Calvert C, Kennedy N, Goodhand J, Ahmad T. Faecal calprotectin effectively excludes inflammatory bowel disease in 789 symptomatic young adults with/without alarm symptoms: a prospective UK primary care cohort study. *Aliment Pharmacol Ther*. 2018;47. doi: 10.1111/apt.14563.
65. Turvill J, Turnock D, Holmes H, Jones A, McLaughlan E, Hilton V, Marriott S. Evaluation of the clinical and cost-effectiveness of the York Faecal Calprotectin Care Pathway. *Frontline Gastroenterol*. 2018;9:285-94. doi: 10.1136/flgastro-2018-100962.
66. Irwig L, Bossuyt P, Glasziou P, Gatsonis C, Lijmer J. Designing studies to ensure that estimates of test accuracy are transferable. *BMJ*. 2002;324(7338):669-71. doi: 10.1136/bmj.324.7338.669.
67. Willis BH. Spectrum bias: why clinicians need to be cautious when applying diagnostic test studies. *Fam Pract*. 2008;25(5):390-6. doi: 10.1093/fampra/cmn051.
68. Leeftang MM, Bossuyt PM, Irwig L. Diagnostic test accuracy may vary with prevalence: implications for evidence-based diagnosis. *J Clin Epidemiol*. 2009;62(1):5-12. doi: 10.1016/j.jclinepi.2008.04.007.
69. Ormston R, Spencer L, Barnard M, Snape D. The foundations of qualitative research. In: Ritchie J, Lewis JD, McNaughton Nicholls C, Ormston R, editors. *Qualitative research practice: a guide for social science students and researchers*. London: SAGE; 2013. p. 1-24.
70. Seale C, Gobo G, Gubrium J, Silverman D. Introduction: inside qualitative research. In: Seale C, Gobo G, Gubrium J, Silverman D, editors. *Qualitative research practice*: SAGE; 2004. p. 1-12.
71. Duncan EAS, Nicol MM. Subtle realism and occupational therapy: an alternative approach to knowledge generation and evaluation. *Br J Occup Ther*. 2004;67(10):453-6. doi: 10.1177/030802260406701006.
72. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. In: Bryman A, Burgess R, editors. *Analyzing qualitative data*. London: Routledge; 1994. p. 172-94.
73. Guba E, Lincoln Y. Competing paradigms in qualitative research. In: Denzin N, Lincoln Y, editors. *Handbook of qualitative research*. Thousand Oaks, CA: SAGE; 1994. p. 105-17.
74. Onwuegbuzie AJ, Johnson RB, Collins KM. Call for mixed analysis: a philosophical framework for combining qualitative and quantitative approaches. *Int J Mult Res Approaches*. 2009;3(2):114-39. doi: 10.5172/mra.3.2.114.

75. Maxwell J. What is realism, and why should qualitative researchers care? A realist approach for qualitative research. Los Angeles, US: SAGE; 2011. p. 3-13.
76. Cohen D, Crabtree B. Qualitative research guidelines project. 2006 [cited 10/02/2020]; Available from: <http://www.qualres.org/>.
77. Lacey A, Luff D. Qualitative research analysis. The NIHR RDS for the East Midlands / Yorkshire & the Humber. 2007 [cited 10/02/2020]; Available from: https://www.rds-yh.nihr.ac.uk/wp-content/uploads/2013/05/9_Qualitative_Data_Analysis_Revision_2009.pdf.
78. Gale NK, Heath G, Cameron E, Rashid S, Redwood S. Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Med Res Methodol*. 2013;13:117. doi: 10.1186/1471-2288-13-117.
79. Angen MJ. Evaluating interpretive inquiry: reviewing the validity debate and opening the dialogue. *Qual Health Res*. 2000;10(3):378-95. doi: 10.1177/104973230001000308.
80. Barnard M. Critical qualitative theory: opening up the black box of criminal justice interventions. In: Bowen E, Brown S, editors. Perspectives on evaluating criminal justice and corrections (advances in program evaluation, Volume 13): Emerald Group Publishing Limited; 2012. p. 129-41.
81. Haynes B. Forming research questions. In: Haynes B, Sackett D, Guyatt G, Tugwell P, editors. Clinical epidemiology: how to do clinical practice research. US: Lippincott, Williams and Wilkins; 2006. p. 3-14.
82. Egger M, Smith G, O'Rourke K. Rationale, potentials, and promise of systematic reviews. In: Egger M, Smith G, Altman D, editors. Systematic reviews in health care: meta-analysis in context: BMJ; 2001. p. 3-22.
83. Ferrante di Ruffano L, Dinnes J, Sitch AJ, Hyde C, Deeks JJ. Test-treatment RCTs are susceptible to bias: a review of the methodological quality of randomized trials that evaluate diagnostic tests. *BMC Med Res Methodol*. 2017;17(1):35. doi: 10.1186/s12874-016-0287-z.
84. Bossuyt PM, Olsen M, Hyde C, Cohen JF. An analysis reveals differences between pragmatic and explanatory diagnostic accuracy studies. *J Clin Epidemiol*. 2020;117:29-35. doi: <https://doi.org/10.1016/j.jclinepi.2019.09.017>.
85. Sedgwick P. Retrospective cohort studies: advantages and disadvantages. *BMJ*. 2014;348:g1072. doi: 10.1136/bmj.g1072.
86. Lis Y, Mann RD. The VAMP research multi-purpose database in the UK. *J Clin Epidemiol*. 1995;48(3):431-43. doi: [https://doi.org/10.1016/0895-4356\(94\)00137-F](https://doi.org/10.1016/0895-4356(94)00137-F).
87. Department of Health, Royal College of General Practitioners, British Medical Association. The good practice guidelines for GP electronic patient records Version 4. 2011 [cited 27/11/2019]; Available from: <https://www.gov.uk/government/publications/the-good-practice-guidelines-for-gp-electronic-patient-records-version-4-2011>.
88. Kontopantelis E, Stevens RJ, Helms PJ, Edwards D, Doran T, Ashcroft DM. Spatial distribution of clinical computer systems in primary care in England in 2016 and implications for primary care electronic medical record databases: a cross-sectional population study. *BMJ Open*. 2018;8(2):e020738. doi: 10.1136/bmjopen-2017-020738.
89. Gentil ML, Cuggia M, Fiquet L, Hagenbourger C, Le Berre T, Banatre A, Renault E, Bouzille G, Chapron A. Factors influencing the development of primary care data collection projects from electronic health records: a systematic review of the literature. *BMC Med Inform Decis Mak*. 2017;17(1):139. doi: 10.1186/s12911-017-0538-x.
90. Vezyridis P, Timmons S. Evolution of primary care databases in UK: a scientometric analysis of research output. *BMJ Open*. 2016;6(10):e012785-e. doi: 10.1136/bmjopen-2016-012785.
91. Lewis JD, Bilker WB, Weinstein RB, Strom BL. The relationship between time since registration and measured incidence rates in the General Practice Research Database. *Pharmacoepidemiol Drug Saf*. 2005;14(7):443-51. doi: 10.1002/pds.1115.
92. Blak BT, Thompson M, Dattani H, Bourke A. Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates. *Inform Prim Care*. 2011;19(4):251-5. doi: 10.14236/jhi.v19i4.820.
93. Bourke A, Dattani H, Robinson M. Feasibility study and methodology to create a quality-evaluated database of primary care data. *Inform Prim Care*. 2004;12(3):171-7. doi: 10.14236/jhi.v12i3.124.
94. IMS Health. THIN data guide for researchers Version 3.0. 2015.
95. Ng SC, Shi HY, Hamidi N, Underwood FE, Tang W, Benchimol EI, Panaccione R, Ghosh S, Wu JCY, Chan FKL, Sung JYJ, Kaplan GG. Worldwide incidence and prevalence

- of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet*. 2018;390(10114):2769-78. doi: 10.1016/s0140-6736(17)32448-0.
96. Hamilton B, Green H, Heerasing N, Hendy P, Moore L, Chanchlani N, Walker G, Bewshea C, Kennedy NA, Ahmad T, Goodhand J. Incidence and prevalence of inflammatory bowel disease in Devon, UK. *Frontline Gastroenterol*. 2020;0:1-10. doi: 10.1136/flgastro-2019-101369.
97. Pasvol TJ, Horsfall L, Bloom S, Segal AW, Sabin C, Field N, Rait G. Incidence and prevalence of inflammatory bowel disease in UK primary care: a population-based cohort study. *BMJ Open*. 2020;10(7):e036584. doi: 10.1136/bmjopen-2019-036584.
98. King D, Reulen RC, Thomas T, Chandan JS, Thayakaran R, Subramanian A, Gokhale K, Bhala N, Nirantharakumar K, Adderley NJ, Trudgill N. Changing patterns in the epidemiology and outcomes of inflammatory bowel disease in the United Kingdom: 2000-2018. *Aliment Pharmacol Ther*. 2020;51(10):922-34. doi: 10.1111/apt.15701.
99. Sperber AD, Dekel R. Irritable bowel syndrome and co-morbid gastrointestinal and extra-gastrointestinal functional syndromes. *J Neurogastroenterol Motil*. 2010;16(2):113-9. doi: 10.5056/jnm.2010.16.2.113.
100. Kozuch PL, Hanauer SB. Treatment of inflammatory bowel disease: a review of medical therapy. *World J Gastroenterol*. 2008;14(3):354-77. doi: 10.3748/wjg.14.354.
101. The University of Manchester. Clinical Codes Repository. [cited 16/04/2020]; Available from: <https://clinicalcodes.rss.mhs.man.ac.uk/>.
102. Harkness EF, Grant L, O'Brien SJ, Chew-Graham CA, Thompson DG. Using read codes to identify patients with irritable bowel syndrome in general practice: a database study. *BMC Fam Pract*. 2013;14:183-. doi: 10.1186/1471-2296-14-183.
103. NHS Business Services Authority. Prescription Cost Analysis (PCA) data. [cited 12/03/2020]; Available from: <https://www.nhsbsa.nhs.uk/prescription-data/dispensing-data/prescription-cost-analysis-pca-data>.
104. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [computer program]. 2017. Available from: <https://www.R-project.org/>.
105. Hulsen T. Create your own Venn diagram. 2003-2008 [cited 13/11/2019]; Available from: <http://www.biovenn.nl/venndiagram.tk/create.php>.
106. Dhaliwal A, Zeino Z, Tomkins C, Cheung M, Nwokolo C, Smith S, Harmston C, Arasaradnam RP. Utility of faecal calprotectin in inflammatory bowel disease (IBD): what cut-offs should we apply? *Frontline Gastroenterol*. 2015;6(1):14-9. doi: 10.1136/flgastro-2013-100420.
107. Poullis A. GP guidelines FC testing: St George's University Hospital: introduction to faecal calprotectin for GPs. 2015 [cited 04/12/2019]; Available from: <https://www.nice.org.uk/guidance/dg11/resources>.
108. Canavan C, West J, Card T. Calculating total health service utilisation and costs from routinely collected electronic health records using the example of patients with irritable bowel syndrome before and after their first gastroenterology appointment. *Pharmacoeconomics*. 2016;34(2):181-94. doi: 10.1007/s40273-015-0339-y.
109. Miller A, Nightingale AL, Sammon CJ, Mahtani KR, Holt TA, McHugh NJ, Luqmani RA. Estimating the diagnostic accuracy of rheumatoid factor in UK primary care: a study using the clinical practice research datalink. *Rheumatology*. 2015;54(10):1882-9. doi: 10.1093/rheumatology/kev131.
110. Sammon CJ, Miller A, Mahtani KR, Holt TA, McHugh NJ, Luqmani RA, Nightingale AL. Missing laboratory test data in electronic general practice records: analysis of rheumatoid factor recording in the clinical practice research datalink. *Pharmacoepidemiol Drug Saf*. 2015;24(5):504-9. doi: 10.1002/pds.3750.
111. McBride D, Hardoon S, Walters K, Gilmour S, Raine R. Explaining variation in referral from primary to secondary care: cohort study. *BMJ*. 2010;341:c6267. doi: 10.1136/bmj.c6267.
112. Gavin DR, Valori RM, Anderson JT, Donnelly MT, Williams JG, Swarbrick ET. The national colonoscopy audit: a nationwide assessment of the quality and safety of colonoscopy in the UK. *Gut*. 2013;62(2):242-9. doi: 10.1136/gutjnl-2011-301848.
113. Garcia Rodriguez LA, Gonzalez-Perez A, Johansson S, Wallander MA. Risk factors for inflammatory bowel disease in the general population. *Aliment Pharmacol Ther*. 2005;22(4):309-15. doi: 10.1111/j.1365-2036.2005.02564.x.

114. Shivananda S, Lennard-Jones J, Logan R, Fear N, Price A, Carpenter L, van Blankenstein M. Incidence of inflammatory bowel disease across Europe: is there a difference between north and south? Results of the European Collaborative Study on Inflammatory Bowel Disease (EC-IBD). *Gut*. 1996;39(5):690-7. doi: 10.1136/gut.39.5.690.
115. Molodecky NA, Soon IS, Rabi DM, Ghali WA, Ferris M, Chernoff G, Benchimol EI, Panaccione R, Ghosh S, Barkema HW, Kaplan GG. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology*. 2012;142(1):46-54. doi: 10.1053/j.gastro.2011.10.001.
116. Vermeire S, Van Assche G, Rutgeerts P. Classification of inflammatory bowel disease: the old and the new. *Curr Opin Gastroenterol*. 2012;28(4):321-6. doi: 10.1097/MOG.0b013e328354be1e.
117. Rogers EM. Diffusion of innovations. 3rd ed. New York Free Press; 1983.
118. CSO Faecal Calprotectin Working Group. Faecal calprotectin in primary care as a decision diagnostic for inflammatory bowel disease and irritable bowel syndrome. 2018 [cited 09/07/2018]; Available from: <https://www.pcc-cic.org.uk/article/updated-guidance-help-gps-easily-diagnose-serious-bowel-conditions>.
119. Ng S, Khan K, Box B. Faecal calprotectin: a good test used badly [abstract]. *Colorectal Dis*. 2017;19(Suppl 4):64. doi: <http://dx.doi.org/10.1111/codi.13851>.
120. Lichtenstein GR, Loftus EV, Isaacs KL, Regueiro MD, Gerson LB, Sands BE. ACG clinical guideline: management of Crohn's disease in adults. *Am J Gastroenterol*. 2018;113:481. doi: 10.1038/ajg.2018.27
<https://www.nature.com/articles/ajg201827#supplementary-information>.
121. The Gastroenterological Society of Australia. Australian guidelines for general practitioners and physicians: inflammatory bowel disease. 2017 [cited 08/10/2018]; Available from: http://cart.gesa.org.au/membes/files/Clinical%20Guidelines%20and%20Updates/2017_IBD_guidelines_DRAFT_for_consultation_20102017.pdf.
122. Preiß J, Bokemeyer B, Buhr H, Dignaß A, Häuser W, Hartmann F, Herrlinger K, Kaltz B, Kienle P, Kruis W, Kucharzik T, Langhorst J, Schreiber S, Siegmund B, Stallmach A, Stange E, Stein J, Hoffmann J. Updated German clinical practice guideline on "Diagnosis and treatment of Crohn's disease". 2014 [cited 08/10/2018]; Available from: https://www.dgvs.de/wp-content/uploads/2016/11/DGVS_Empfehlung_fuer_Diagnostik_und_Therapie_des_Morbus_Crohn.pdf.
123. Gomollon F, Dignass A, Annese V, Tilg H, Van Assche G, Lindsay JO, Peyrin-Biroulet L, Cullen GJ, Daperno M, Kucharzik T, Rieder F, Almer S, Armuzzi A, Harbord M, Langhorst J, Sans M, Chowers Y, Fiorino G, Juillerat P, Mantzaris GJ, Rizzello F, Vavricka S, Gionchetti P. 3rd European evidence-based consensus on the diagnosis and management of crohn's disease 2016: part 1: diagnosis and medical management. *J Crohns Colitis*. 2017;11(1):3-25. doi: 10.1093/ecco-jcc/jjw168.
124. Turvill J. High negative predictive value of a normal faecal calprotectin in patients with symptomatic intestinal disease. *Frontline Gastroenterol*. 2012;3(1):21-8. doi: 10.1136/flgastro-2011-100011.
125. Dolwani S, Metzner M, Wassell JJ, Yong A, Hawthorne AB. Diagnostic accuracy of faecal calprotectin estimation in prediction of abnormal small bowel radiology. *Aliment Pharmacol Ther*. 2004;20(6):615-21. doi: 10.1111/j.1365-2036.2004.02128.x.
126. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, Leeflang MM, Sterne JA, Bossuyt PM. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-36. doi: 10.7326/0003-4819-155-8-201110180-00009.
127. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58(10):982-90. doi: 10.1016/j.jclinepi.2005.02.022.
128. Partlett C, Takwoingi Y. Meta-analysis of test accuracy studies in R: a summary of user-written programs and step-by-step guide to using glmer. Version 1.0. 2016 30/01/2018 [cited 30/01/2018]; Available from: <http://methods.cochrane.org/sdt/>.
129. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, Clifford T, Cohen JF, Deeks JJ, Gatsonis C, Hooft L, Hunt HA, Hyde CJ, Korevaar DA, Leeflang MMG, Macaskill P, Reitsma JB, Rodin R, Rutjes AWS, Salameh JP, Stevens A, Takwoingi Y,

- Tonelli M, Weeks L, Whiting P, Willis BH. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA*. 2018;319(4):388-96. doi: 10.1001/jama.2017.19163.
130. Oyaert M, Boel A, Jacobs J, Van den Bremt S, De Sloovere M, Vanpoucke H, Van Hoovels L. Analytical performance and diagnostic accuracy of six different faecal calprotectin assays in inflammatory bowel disease. *Clin Chem Lab Med*. 2017;21:21. doi: <https://dx.doi.org/10.1515/cclm-2016-1012>.
131. Labaere D, Smismans A, Van Olmen A, Christiaens P, D'Haens G, Moons V, Cuyle PJ, Frans J, Bossuyt P. Comparison of six different calprotectin assays for the assessment of inflammatory bowel disease. *United European Gastroenterol J*. 2014;2(1):30-7. doi: <https://dx.doi.org/10.1177/2050640613518201>.
132. Schoepfer AM, Trummler M, Seeholzer P, Seibold-Schmid B, Seibold F. Discriminating IBD from IBS: comparison of the test performance of fecal markers, blood leukocytes, CRP, and IBD antibodies. *Inflamm Bowel Dis*. 2008;14(1):32-9. doi: 10.1002/ibd.20275.
133. Otten CM, Kok L, Witteman BJ, Baumgarten R, Kampman E, Moons KG, de Wit NJ. Diagnostic performance of rapid tests for detection of fecal calprotectin and lactoferrin and their ability to discriminate inflammatory from irritable bowel syndrome. *Clin Chem Lab Med*. 2008;46(9):1275-80. doi: 10.1515/cclm.2008.246.
134. El-Badry A, Sedrak H, Rashed L. Faecal calprotectin in differentiating between functional and organic bowel diseases. *Arab J Gastroenterol*. 2010;11(2):70-3. doi: <https://doi.org/10.1016/j.ajg.2010.04.009>.
135. Burri E, Manz M, Rothen C, Rossi L, Beglinger C, Lehmann FS. Monoclonal antibody testing for fecal calprotectin is superior to polyclonal testing of fecal calprotectin and lactoferrin to identify organic intestinal disease in patients with abdominal discomfort. *Clin Chim Acta*. 2013;416:41-7. doi: 10.1016/j.cca.2012.11.008.
136. Damms A, Bischoff SC. Validation and clinical significance of a new calprotectin rapid test for the diagnosis of gastrointestinal diseases. *Int J Colorectal Dis*. 2008;23(10):985-92. doi: 10.1007/s00384-008-0506-0.
137. Mowat C, Digby J, Strachan JA, Wilson R, Carey FA, Fraser CG, Steele RJ. Faecal haemoglobin and faecal calprotectin as indicators of bowel disease in patients presenting to primary care with bowel symptoms. *Gut*. 2016;65(9):1463-9. doi: <https://dx.doi.org/10.1136/gutjnl-2015-309579>.
138. De Sloovere MM, De Smet D, Baert FJ, Debrabandere J, Vanpoucke HJ. Analytical and diagnostic performance of two automated fecal calprotectin immunoassays for detection of inflammatory bowel disease. *Clin Chem Lab Med*. 2017;55(9):1435-46. doi: <https://dx.doi.org/10.1515/cclm-2016-0796>.
139. Livzan MA, Lyalukova EA, Nechaeva G, Osipenko MF, Dolgih TI. [Fecal noninvasive tests (calprotectin, transferrin, hemoglobin) in complex diagnosis of diseases of intestines]. *Eksperimental'Naia i Klinicheskaia Gastroenterologija*. 2015(3):34-8.
140. Oyaert M, Trouve C, Baert F, De Smet D, Langlois M, Vanpoucke H. Comparison of two immunoassays for measurement of faecal calprotectin in detection of inflammatory bowel disease: (pre)-analytical and diagnostic performance characteristics. *Clin Chem Lab Med*. 2014;52(3):391-7. doi: <https://dx.doi.org/10.1515/cclm-2013-0699>.
141. Sostres C, Lue A, Barra Pardos MV, Hijos G, Perales A, Puente JJ, Lanás A, Gomollon F. Combination of FOBT and fecal calprotectin may be useful for reducing unnecessary colonoscopies in symptomatic patients [abstract]. *United European Gastroenterol J*. 2017;5(Suppl 5):A779. doi: <http://dx.doi.org/10.1177/2050640617725676>.
142. Alrubaiy L, Malik A, Rees I, Bowen D. Usefulness of fecal calprotectin in clinical practice in a district general hospital [abstract]. *Inflamm Bowel Dis*. 2012;18(Suppl 1):S53-4. doi: <http://dx.doi.org/10.1002/ibd.23058>.
143. Carroccio A, Iacono G, Cottone M, Di Prima L, Cartabellotta F, Cavataio F, Scalici C, Montalto G, Di Fede G, Rini G, Notarbartolo A, Averna MR. Diagnostic accuracy of fecal calprotectin assay in distinguishing organic causes of chronic diarrhea from irritable bowel syndrome: a prospective study in adults and children. *Clin Chem*. 2003;49(6):861-7. doi: <https://doi.org/10.1373/49.6.861>.
144. Caviglia GP, Pantaleoni S, Touscoz GA, Adriani A, Rosso C, Smedile A, Pellicano R, Astegiano M, Bresso F. Fecal calprotectin is an effective diagnostic tool that differentiates inflammatory from functional intestinal disorders. *Scand J Gastroenterol*. 2014;49(12):1419-24. doi: <https://dx.doi.org/10.3109/00365521.2014.934913>.

145. Tibble JA, Sigthorsson G, Foster R, Forgacs I, Bjarnason I. Use of surrogate markers of inflammation and Rome criteria to distinguish organic from nonorganic intestinal disease. *Gastroenterology*. 2002;123(2):450-60. doi: 10.1053/gast.2002.34755.
146. Tan M, Subramaniam Y, Cheong C, Hasan F, Moran GW, Fateen W, Jawhari A. The role of faecal calprotectin in a primary care referral pathway in the UK [abstract]. *United European Gastroenterol J*. 2016;3(Suppl 5):A433. doi: <http://dx.doi.org/10.1177/2050640616663689>.
147. Limburg PJ, Ahlquist DA, Sandborn WJ, Mahoney DW, Devens ME, Harrington JJ, Zinsmeister AR. Fecal calprotectin levels predict colorectal inflammation among patients with chronic diarrhea referred for colonoscopy. *Am J Gastroenterol*. 2000;95(10):2831-7. doi: 10.1111/j.1572-0241.2000.03194.x.
148. Rosenfeld G, Greenup AJ, Round A, Takach O, Halparin L, Saadeddin A, Ho JK, Lee T, Enns R, Bressler B. FOCUS: future of fecal calprotectin utility study in inflammatory bowel disease. *World J Gastroenterol*. 2016;22(36):8211-8. doi: <https://dx.doi.org/10.3748/wjg.v22.i36.8211>.
149. Oostenbrink R, Moons KGM, Bleeker SE, Moll HA, Grobbee DE. Diagnostic research on routine care data: prospects and problems. *J Clin Epidemiol*. 2003;56(6):501-6. doi: [https://doi.org/10.1016/S0895-4356\(03\)00080-5](https://doi.org/10.1016/S0895-4356(03)00080-5).
150. Holtman GA, Berger MY, Burger H, Deeks JJ, Donner-Banzhoff N, Fanshawe TR, Koshiaris C, Leeflang MM, Oke JL, Perera R, Reitsma JB, Van den Bruel A. Development of practical recommendations for diagnostic accuracy studies in low-prevalence situations. *J Clin Epidemiol*. 2019;114:38-48. doi: <https://doi.org/10.1016/j.jclinepi.2019.05.018>.
151. van der Schouw YT, Van Dijk R, Verbeek AL. Problems in selecting the adequate patient population from existing data files for assessment studies of new diagnostic tests. *J Clin Epidemiol*. 1995;48(3):417-22. doi: 10.1016/0895-4356(94)00144-f.
152. Zwetsloot-Schonk JH. Use of a hospital database to determine the characteristics of diagnostic tests. *Methods Inf Med*. 1990;29(3):205-12.
153. Hlatky MA, Pryor DB, Harrell FE, Jr., Califf RM, Mark DB, Rosati RA. Factors affecting sensitivity and specificity of exercise electrocardiography. Multivariable analysis. *Am J Med*. 1984;77(1):64-71. doi: 10.1016/0002-9343(84)90437-6.
154. Begg CB. Biases in the assessment of diagnostic tests. *Stat Med*. 1987;6(4):411-23. doi: 10.1002/sim.4780060402.
155. Diamond GA. Selection bias and the evaluation of diagnostic tests: a meta-dissent. *J Chron Dis*. 1986;39(5):359-60.
156. Ramraj R, Garcia A, Mosen D, Waiwaiole L, Smith N. Utility of fecal calprotectin in evaluation of chronic gastrointestinal symptoms in primary care. *Clin Pediatr (Phila)*. 2018;57(9):1058-63. doi: 10.1177/0009922817744607.
157. 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-e37. doi: 10.3399/bjgp12X616346.
158. Clark-Snustad KD, Lee SD. Diagnosing inflammatory bowel disease and differentiating it from potential mimics. *Techniques in Gastrointestinal Endoscopy*. 2016;18(3):108-15. doi: <https://doi.org/10.1016/j.tgie.2016.07.006>.
159. Lamb CA, Kennedy NA, Raine T, Hendy PA, Smith PJ, Limdi JK, Hayee B, Lomer MCE, Parkes GC, Selinger C, Barrett KJ, Davies RJ, Bennett C, Gittens S, Dunlop MG, Faiz O, Fraser A, Garrick V, Johnston PD, Parkes M, Sanderson J, Terry H, Gaya DR, Iqbal TH, Taylor SA, Smith M, Brookes M, Hansen R, Hawthorne AB. British Society of Gastroenterology consensus guidelines on the management of inflammatory bowel disease in adults. *Gut*. 2019;68(Suppl 3):s1-s106. doi: 10.1136/gutjnl-2019-318484.
160. Venkatraman ES. A permutation test to compare receiver operating characteristic curves. *Biometrics*. 2000;56(4):1134-8. doi: 10.1111/j.0006-341x.2000.01134.x.
161. Holtman GA. Diagnostic strategies in children with chronic gastrointestinal symptoms in primary care [thesis]. Rijksuniversiteit Groningen 2016 [cited 26/09/2017]; Available from: [http://www.rug.nl/research/portal/publications/diagnostic-strategies-in-children-with-chronic-gastrointestinal-symptoms-in-primary-care\(f36e28e3-d8ce-4eaa-848d-af7e9dadcd185\).html](http://www.rug.nl/research/portal/publications/diagnostic-strategies-in-children-with-chronic-gastrointestinal-symptoms-in-primary-care(f36e28e3-d8ce-4eaa-848d-af7e9dadcd185).html).
162. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, Wood AM, Carpenter JR. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009;338:b2393. doi: 10.1136/bmj.b2393.

163. Rutjes AW, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PM. Evaluation of diagnostic tests when there is no gold standard: a review of methods. *Health Technol Assess.* 2007;11(50):iii, ix-51.
164. de Groot JAH, Dendukuri N, Janssen KJM, Reitsma JB, Bossuyt PM, Moons KGM. Adjusting for differential verification bias in diagnostic accuracy studies: a Bayesian approach. *Epidemiology.* 2011;22(2):234-41. doi: 10.1097/EDE.0b013e318207fc5c.
165. Willis BH, Hyde CJ. Estimating a test's accuracy using tailored meta-analysis: how setting-specific data may aid study selection. *J Clin Epidemiol.* 2014;67(5):538-46. doi: 10.1016/j.jclinepi.2013.10.016.
166. Willis BH, Hyde CJ. What is the test's accuracy in my practice population? Tailored meta-analysis provides a plausible estimate. *J Clin Epidemiol.* 2015;68(8):847-54. doi: 10.1016/j.jclinepi.2014.10.002.
167. Willis BH, Coomar D, Baragilly M. Tailored meta-analysis: an investigation of the correlation between the test positive rate and prevalence. *J Clin Epidemiol.* 2019;106:1-9. doi: <https://doi.org/10.1016/j.jclinepi.2018.09.013>.
168. Sharbatdaran M, Halakou A, Kashifard M, Bijani A, Firozjaei A, Hossaini A, Siadati S. Fecal calprotectin level in patients with IBD and noninflammatory disease of colon: a study in Babol, Northern Iran. *Caspian J Intern Med.* 2018;9(1):60-4. doi: 10.22088/cjim.9.1.60.
169. Fries W, Belvedere A, Alecci U, Alibrandi A, Bruno L, Crescenti A, Di Geronimo L, Inferrera S, Marino S, Scoglio R. Early diagnosis in inflammatory bowel disease (IBD): a pilot study [abstract]. *Dig Liver Dis.* 2018;50 (Suppl 1):e93-e4. doi: 10.1016/S1590-8658(18)30325-6.
170. Riley RD, Ahmed I, Debray TPA, Willis BH, Noordzij JP, Higgins JPT, Deeks JJ. Summarising and validating test accuracy results across multiple studies for use in clinical practice. *Stat Med.* 2015;34(13):2081-103. doi: 10.1002/sim.6471.
171. Reyna VF. Theories of medical decision making and health: an evidence-based approach. *Med Decis Making.* 2008;28(6):829-33. doi: 10.1177/0272989x08327069.
172. Foot C, Naylor C, Imison C. The quality of diagnosis and referral. The King's Fund; 2010 [cited 30/01/2020]; Available from: <https://www.kingsfund.org.uk/projects/gp-inquiry/diagnosis-referral>.
173. Sullivan CO, Omar RZ, Ambler G, Majeed A. Case-mix and variation in specialist referrals in general practice. *Br J Gen Pract.* 2005;55(516):529-33.
174. O'Donnell CA. Variation in GP referral rates: what can we learn from the literature? *Fam Pract.* 2000;17(6):462-71. doi: 10.1093/fampra/17.6.462.
175. Wilkin D, Smith A. Explaining variation in general practitioner referrals to hospital. *Fam Pract.* 1987;4(3):160-9. doi: 10.1093/fampra/4.3.160.
176. Newton J, Hayes VIC, Hutchinson A. Factors influencing general practitioners' referral decisions. *Fam Pract.* 1991;8(4):308-13. doi: 10.1093/fampra/8.4.308.
177. Morrell DC, Gage HG, Robinson NA. Referral to hospital by general practitioners. *J R Coll Gen Pract.* 1971;21(103):77-85.
178. Dowie R. General practitioners and consultants: a study of outpatient referrals. King Edward's Hospital Fund for London: Oxford University Press 1983. 208 p.
179. Coyle E, Hanley K, Sheerin J. Who goes where? A prospective study of referral patterns within a newly established primary care team. *Ir J Med Sci.* 2011;180(4):845-9. doi: 10.1007/s11845-011-0724-2.
180. Ringberg U, Fleten N, Forde OH. Examining the variation in GPs' referral practice: a cross-sectional study of GPs' reasons for referral. *Br J Gen Pract.* 2014;64(624):e426-33. doi: 10.3399/bjgp14X680521.
181. Harris M, Vedsted P, Esteva M, Murchie P, Aubin-Auger I, Azuri J, Brekke M, Buczkowski K, Buono N, Costiug E, Dinant G-J, Foreva G, Gašparović Babić S, Hoffman R, Jakob E, Koskela TH, Marzo-Castillejo M, Neves AL, Petek D, Petek Ster M, Sawicka-Powierza J, Schneider A, Smyrnakis E, Streit S, Thulesius H, Weltermann B, Taylor G. Identifying important health system factors that influence primary care practitioners' referrals for cancer suspicion: a European cross-sectional survey. *BMJ Open.* 2018;8(9):e022904. doi: 10.1136/bmjopen-2018-022904.
182. Calnan M, Payne S, Kemple T, Rosedale M, Ingram J. A qualitative study exploring variations in GPs' out-of-hours referrals to hospital. *Br J Gen Pract.* 2007;57(542):706-13.
183. Nandy S, Chalmers-Watson C, Gantley M, Underwood M. Referral for minor mental illness: a qualitative study. *Br J Gen Pract.* 2001;51(467):461-5.

184. Bowling A, Redfern J. The process of outpatient referral and care: the experiences and views of patients, their general practitioners, and specialists. *Br J Gen Pract*. 2000;50(451):116-20.
185. Elstein AS, Schwartz A. Clinical problem solving and diagnostic decision making: selective review of the cognitive literature. *BMJ*. 2002;324(7339):729-32. doi: 10.1136/bmj.324.7339.729.
186. Tversky A, Kahneman D. Judgment under Uncertainty: Heuristics and Biases. *Science*. 1974;185(4157):1124-31. doi: 10.1126/science.185.4157.1124.
187. Hamm RM. Clinical intuition and clinical analysis: expertise and the cognitive continuum. In: Dowe J, Elstein A, editors. *Professional judgement: a reader in clinical decision making*. Cambridge: Cambridge University Press; 1988. p. 78-109.
188. Kushniruk AW. Analysis of complex decision-making processes in health care: cognitive approaches to health informatics. *J Biomed Inform*. 2001;34(5):365-76. doi: 10.1006/jbin.2001.1021.
189. King N, Bailey J, Newton P. Analysing general practitioners' referral decisions. I. Developing an analytical framework. *Fam Pract*. 1994;11(1):3-8. doi: <https://doi.org/10.1093/fampra/11.1.3>.
190. Mason M. Sample size and saturation in PhD studies using qualitative interviews. 2010 17/12/2015 [cited 17/12/2015]; Available from: <http://www.qualitative-research.net/index.php/fqs/article/view/1428/3027>.
191. Guest G, Namey EE, Mitchell ML. *Collecting qualitative data: a field manual for applied research*. London, UK: Sage; 2013.
192. Hayes H, Buckland S, Tarpey M, of the INVOLVE Coordinating Centre. Briefing notes for researchers: public involvement in NHS, public health and social care research. 2012 [cited 23/04/2020]; Available from: <https://www.invo.org.uk/resource-centre/resource-for-researchers/>.
193. Cummins RO, Jarman B, White PM. Do general practitioners have different "referral thresholds"? *BMJ*. 1981;282(6269):1037-9. doi: 10.1136/bmj.282.6269.1037.
194. Bourne T, De Cock B, Wynants L, Peters M, Van Audenhove C, Timmerman D, Van Calster B, Jalnibrant M. Doctors' perception of support and the processes involved in complaints investigations and how these relate to welfare and defensive practice: a cross-sectional survey of the UK physicians. *BMJ Open*. 2017;7(11):e017856. doi: 10.1136/bmjopen-2017-017856.
195. Jenkins RM. Quality of general practitioner referrals to outpatient departments: assessment by specialists and a general practitioner. *Br J Gen Pract*. 1993;43(368):111-3.
196. Elwyn GJ, Stott NCH. Avoidable referrals? Analysis of 170 consecutive referrals to secondary care. *BMJ*. 1994;309(6954):576-8. doi: 10.1136/bmj.309.6954.576.
197. Fertig A, Roland M, King H, Moore T. Understanding variation in rates of referral among general practitioners: are inappropriate referrals important and would guidelines help to reduce rates? *BMJ*. 1993;307(6917):1467-70. doi: 10.1136/bmj.307.6917.1467.
198. Dixon A, Robertson R, Bal R. The experience of implementing choice at point of referral: a comparison of the Netherlands and England. *Health Econ Policy Law*. 2010;5(3):295-317. doi: 10.1017/S1744133110000058.
199. Armstrong D, Fry J, Armstrong P. Doctors' perceptions of pressure from patients for referral. *BMJ*. 1991;302(6786):1186-8. doi: 10.1136/bmj.302.6786.1186
200. Thorsen O, Hartveit M, Johannessen JO, Fosse L, Eide GE, Schulz J, Baerheim A. Typologies in GPs' referral practice. *BMC Fam Pract*. 2016;17:76. doi: 10.1186/s12875-016-0495-y.
201. Bailey J, King N, Newton P. Analysing general practitioners' referral decisions II. Applying the analytical framework: do high and low referrers differ in factors influencing their referral decisions? *Fam Pract*. 1994;11(1):9-14. doi: 10.1093/fampra/11.1.9.
202. Blank L, Baxter S, Woods HB, Goyder E, Lee A, Payne N, Rimmer M. What is the evidence on interventions to manage referral from primary to specialist non-emergency care? A systematic review and logic model synthesis. *Health Services and Delivery Research, No 324*. 2015:1-429. doi: 10.3310/hsdr03240.
203. Kahneman D. *Thinking, fast and slow*. London, England: Penguin Books; 2012.
204. Harbison J. Clinical decision making in nursing: theoretical perspectives and their relevance to practice. *J Adv Nurs*. 2001;35(1):126-33. doi: 10.1046/j.1365-2648.2001.01816.x.

205. Banning M. A review of clinical decision making: models and current research. *J Clin Nurs*. 2008;17(2):187-95. doi: 10.1111/j.1365-2702.2006.01791.x.
206. Truglio-Londrigan M, Slyer JT, Singleton JK, Worrall P. A qualitative systematic review of internal and external influences on shared decision-making in all health care settings. *JBI Libr Syst Rev*. 2012;10(58):4633-46. doi: 10.11124/jbisir-2012-432.
207. Elliott N. 'Mutual intacting': a grounded theory study of clinical judgement practice issues. *J Adv Nurs*. 2010;66(12):2711-21. doi: 10.1111/j.1365-2648.2010.05412.x.
208. Scott IC, Mangat N, MacGregor A, Raza K, Mallen CD, Hider SL. Primary care challenges in diagnosing and referring patients with suspected rheumatoid arthritis: a national cross-sectional GP survey. *Rheumatol Adv Pract*. 2018;2(1):1-6. doi: 10.1093/rap/rky012.
209. Lincoln Y, Guba E. *Naturalistic Inquiry*. Newbury Park, CA: SAGE Publications, Inc.; 1985.
210. Knottnerus AJ, Leffers P. The influence of referral patterns on the characteristics of diagnostic tests. *J Clin Epidemiol*. 1992;45(10):1143-54. doi: [https://doi.org/10.1016/0895-4356\(92\)90155-G](https://doi.org/10.1016/0895-4356(92)90155-G).
211. Willis BH. Empirical evidence that disease prevalence may affect the performance of diagnostic tests with an implicit threshold: a cross-sectional study. *BMJ Open*. 2012;2(1):e000746-e. doi: 10.1136/bmjopen-2011-000746.
212. Turvill J, Turnock D. Audit of the impact of the York faecal calprotectin care pathway on colonoscopy activity. *Frontline Gastroenterol*. 2019;0:1-5. doi: 10.1136/flgastro-2019-101315.
213. NHS Digital UK. Read Codes. 2017 [cited 25/10/2017]; Available from: <https://digital.nhs.uk/article/1104/Read-Codes>.
214. BNF Publications. [cited 12/03/2017]; Available from: <https://www.bnf.org/>.
215. Mulder DJ, Noble AJ, Justinich CJ, Duffin JM. A tale of two diseases: the history of inflammatory bowel disease. *J Crohns Colitis*. 2014;8(5):341-8. doi: <https://doi.org/10.1016/j.crohns.2013.09.009>.
216. Bharathi S, Moncur P, Holbrook I, Kelly S. A normal faecal calprotectin has a high negative predictive value in cases of suspected irritable bowel syndrome [abstract]. *Gastroenterology*. 2005;128(Suppl 2):A459. doi: <https://doi.org/10.1053/j.gastro.2005.04.003>.
217. Garcia Sanchez M, Gonzalez R, Iglesias Flores E, Gomez Camacho F, Casais Juanena L, Cerezo Ruiz A, Montero Perez-Barquero M, Muntane J, de Dios Vega JF. [Diagnostic value of fecal calprotectin in predicting an abnormal colonoscopy]. *Med Clin (Barc)*. 2006;127(2):41-6.
218. Lee S, Borthwick H, Dhar A. Faecal calprotectin testing in primary and secondary care: are the current manufacturer's cut-off values clinically useful? [abstract]. *J Crohns Colitis*. 2014;8(Suppl 1):S155-S6. doi: [https://doi.org/10.1016/S1873-9946\(14\)60344-5](https://doi.org/10.1016/S1873-9946(14)60344-5)
219. Licata A, Randazzo C, Cappello M, Calvaruso V, Butera G, Florena AM, Peralta S, Camma C, Craxi A. Faecal calprotectin in clinical practice: a noninvasive screening tool for patients with chronic diarrhea. *J Clin Gastroenterol*. 2012;46(6):504-8. doi: 10.1097/MCG.0b013e318248f289.
220. Shitrit AB, Braverman D, Stankiewics H, Shitrit D, Peled N, Paz K. Faecal calprotectin as a predictor of abnormal colonic histology. *Dis Colon Rectum*. 2007;50(12):2188-93. doi: 10.1007/s10350-007-9038-x.

Appendix

Appendix A Use of FC testing in primary care

A1 QUALITY INDICATORS OF THIN DATA FOR RESEARCH PURPOSES

Following the move to a (new) computer system it takes some time before all historic patient records are fully entered into the new system and all patient data have been fully moved. This creates a period of incomplete data and unreliable recording not fit for research purposes.⁹⁴ To address this, IQVIA have assigned three quality indicator dates to each practice that signed up to THIN (Box A1). These are the acceptable mortality reporting date, the computerisation date and the Vision installation date.⁹⁴

Box A1 Definitions of THIN quality indicators

Acceptable mortality reporting (AMR) date

“The AMR date is the year from which the practice is deemed to be reporting all-cause mortality based on predicted numbers of deaths derived from national statistics given the practice’s demographics” (THIN data guide for researchers,⁹⁴ p.74) and should be used to define inclusion criteria for practices’ eligibility to contribute patients to a study. It is a quality indicator of accurate recording for each practice in the THIN database.

Computerisation date

The computerisation date uses information on prescribing and indicates the date when practices have fully moved to electronic health care records defined as the date when the practice started issuing prescriptions using the computer system every day for three consecutive months.

Vision installation date

The Vision installation date is the date when the practice moved to the Vision software.

A2 BRIEF OVERVIEW OF READ CODES

“Read Codes are a coded thesaurus of clinical terms” allowing a standardised way of recording patient findings and procedures.²¹³ There are around 100,000 alphanumeric Read codes available to record clinical information. Read codes are organised hierarchical in chapters in which higher order chapters describe broad disease categories which are subdivided into increasingly specific conditions, for instance chapter J...00 ‘Digestive system diseases’ includes chapter J4...00

'Noninfective enteritis and colitis'. Different Read codes with varying specificity can be used to record a single condition. In order to identify and include all patients with a condition of interest into the study dataset all possible Read codes need to be considered.

A3 CREATING CODE LISTS FOR STUDY VARIABLES

Symptoms and conditions - codes

For conditions and symptoms I created lists of Read codes by interrogating existing code lists (except for functional dyspepsia and Enteropathic arthritis for which no existing Read code lists were available).¹⁰¹ I identified the main Read code chapter(s) from the existing lists and included all Read codes within these for review against my variable definitions. Additional codes were identified by searching the Read code dictionary (a Stata file) provided by IQVIA together with the THIN data guide for researchers.⁹⁴ The Read code dictionary includes the Read code (medcode) made up of up to seven characters and a short description. The dictionary can be searched by Read code or by using words / phrases which match the short description. The aim was to identify all Read codes that can be used to describe a condition or symptom which is a variable of interest in the study. I reviewed the initial Read code lists noting reasons for inclusion and exclusion based on the agreed definitions of conditions and symptoms. My decisions were reviewed by a general practitioner and we resolved discrepancies by discussion. This process is explained for IBD in Box A2.

Box A2 Creating Read code lists for IBD

I designed a search for IBD with the aim to identify all patients who have a definitive diagnosis of IBD. All Read codes in chapter J4...00 'Noninfective enteritis and colitis' were included for review (J4...00 indicates a seven digit alphanumeric Read code where the "... " denotes wildcards of sub-chapters). Supplementary searches for relevant codes outside chapter J4...00 were undertaken using the following terms:

- Regional enteritis
- Crohn's disease (Crohn's and Crohn)
- Colitis
- Proctitis
- Ileitis

Chapter J4...00 contributed 120 Read codes for review. Of the additional searches, Crohn's disease, colitis and proctitis identified 7, 23 and 4 additional Read codes, respectively. These 34 additional codes were added to the list of Read codes for review. The total number of Read codes reviewed was 154. Following review I included a total of 48 Read codes that can be used to record a diagnosis of IBD. The final lists of Read codes are available in Table A1 and were used to extract relevant patient data from the THIN database.

Investigations - codes

All six Read codes available to GPs to record an FC test were used to identify patients with an FC test record.

For the development of a code list for invasive investigations I searched the Read code dictionary and Additional Health Data dictionary for all descriptions ending with 'oscopy' and screened the results for terms relating to investigations of the intestinal tract and the relevant codes. In addition I screened the chapters 'Referral for further care' and 'diagnostic procedures' for additional Read codes.

Final code lists are provided in Table A1.

Drugs - codes

To create exhaustive drug code lists I first identified all generic and proprietary names of IBD specific medications using the Prescription Cost Analysis¹⁰³ data which hold information about all dispensed drugs in England per month. The data

are freely available and are categorised based on the therapeutic groupings used in the BNF.²¹⁴ The data include all generic names with all relevant brand names (used more than 50 times in a month so very rarely used drugs are not included) and is easily searchable by generic name.

To identify withdrawn drugs I compared the 2016 data with the 2010 and 2004 data. This was sufficient for my purpose of identifying patients on IBD treatment as withdrawn drugs will be replaced with an alternative to continue treatment.

I checked my list against the BNF 2016/17.²¹⁴ No further drug names were added to the list.

With the final list of drug names I queried the THIN's drug code dictionary (a complete listing of all drug codes) using all generic and proprietary names identified for each drug to collate all codes available to GPs to record a prescribed IBD drug in the patient records.

In addition to the drug code dictionary I checked the Read code dictionary for Read codes containing the drug names. As the drug name was intended to identify patients on IBD specific medication I was also interested to find Read codes that would identify patients on IBD treatment or with an adverse reaction to IBD treatment.

I used the agreed definitions to exclude irrelevant codes (e.g. eye drops) and noted reasons for inclusions and exclusion. My decisions were reviewed by a general practitioner and we resolved disagreements through discussion. The final lists of drug codes are available in Table A1 which were used to search the THIN database to identify patients with a prescription record for one of the included IBD specific drugs.

Table A1 Final code lists for study variables

Variable	Medcode/drug code/ADH code	Description
IBD	14C4.11	H/O: ulcerative colitis
	J08z900	Orofacial Crohn's disease
	J4...12	Inflammatory bowel disease
	J40..00	Regional enteritis - Crohn's disease
	J40..11	Crohn's disease

	J40..12	Granulomatous enteritis
	J400.00	Regional enteritis of the small bowel
	J400000	Regional enteritis of the duodenum
	J400100	Regional enteritis of the jejunum
	J400200	Crohn's disease of the terminal ileum
	J400300	Crohn's disease of the ileum unspecified
	J400400	Crohn's disease of the ileum NOS
	J400500	Exacerbation of Crohn's disease of small intestine
	J400z00	Crohn's disease of the small bowel NOS
	J401.00	Regional enteritis of the large bowel
	J401000	Regional enteritis of the colon
	J401100	Regional enteritis of the rectum
	J401200	Exacerbation of Crohn's disease of large intestine
	J401z00	Crohn's disease of the large bowel NOS
	J401z11	Crohn's colitis
	J402.00	Regional ileocolitis
	J40z.11	Crohn's disease NOS
	J41..12	Ulcerative colitis and/or proctitis
	J410.00	Ulcerative proctocolitis
	J410000	Ulcerative ileocolitis
	J410100	Ulcerative colitis
	J410200	Ulcerative rectosigmoiditis
	J410300	Ulcerative proctitis
	J410400	Exacerbation of ulcerative colitis
	J410z00	Ulcerative proctocolitis NOS
	J411.00	Ulcerative (chronic) enterocolitis
	J412.00	Ulcerative (chronic) ileocolitis
	J413.00	Ulcerative pancolitis
	J436.00	Microscopic colitis
	J436000	Collagenous colitis
	J436100	Lymphocytic colitis
	J438.00	Left sided colitis
	J4z3.00	Non-infective colitis NOS
	J4z4.00	Non-infective sigmoiditis NOS
	J4z6.00	Indeterminate colitis
	Jyu4000	[X]Other Crohn's disease
	Jyu4100	[X]Other ulcerative colitis
	N031000	Arthropathy in ulcerative colitis
	N031100	Arthropathy in Crohn's disease
	N045300	Juvenile arthritis in Crohn's disease
	N045400	Juvenile arthritis in ulcerative colitis
	ZR3S.00	Crohn's disease activity index
	ZR3S.11	CDAI - Crohn's disease activity index
IBS	14CF.00	History of irritable bowel syndrome
	8CA4a00	Education about FODMAP exclusion diet
	8CA4Z00	Dietary education for irritable bowel syndrome

	8Cm..00	Management of irritable bowel syndrome
	Eu45324	[X]Psychogenic IBS
	J521.00	Irritable colon - Irritable bowel syndrome
	J521.11	Irritable bowel syndrome
	J521.13	Spastic colon
	J521000	Irritable bowel syndrome with diarrhoea
	J521100	Irritable bowel syndrome characterised by constipation
	J521200	IBS characterised by alternating bowel habit
Colorectal cancer	68W2400	Bowel scope (flexible sigmoidoscopy) screen: cancer detected
	68W2500	Bowel scope (flexi-sig) screen: suspected cancer detected
	9Ow1.00	Bowel cancer detected by national screening programme
	B13..00	Malignant neoplasm of colon
	B130.00	Malignant neoplasm of hepatic flexure of colon
	B131.00	Malignant neoplasm of transverse colon
	B132.00	Malignant neoplasm of descending colon
	B133.00	Malignant neoplasm of sigmoid colon
	B134.00	Malignant neoplasm of caecum
	B134.11	Carcinoma of caecum
	B136.00	Malignant neoplasm of ascending colon
	B137.00	Malignant neoplasm of splenic flexure of colon
	B138.00	Malignant neoplasm, overlapping lesion of colon
	B139.00	Hereditary nonpolyposis colon cancer
	B13y.00	Malignant neoplasm of other specified sites of colon
	B13z.00	Malignant neoplasm of colon NOS
	B13z.11	Colonic cancer
	B14..00	Malignant neoplasm of rectum, rectosigmoid junction and anus
	B140.00	Malignant neoplasm of rectosigmoid junction
	B141.00	Malignant neoplasm of rectum
	B141.11	Carcinoma of rectum
	B141.12	Rectal carcinoma
	B142000	Malignant neoplasm of cloacogenic zone
	B14y.00	Malig neop other site rectum, rectosigmoid junction and anus
	B14z.00	Malignant neoplasm rectum,rectosigmoid junction and anus NOS
	B1z0.11	Cancer of bowel
	B575.00	Secondary malignant neoplasm of large intestine and rectum
	B575000	Secondary malignant neoplasm of colon
	B575100	Secondary malignant neoplasm of rectum
	B575z00	Secondary malig neop of large intestine or rectum NOS
	B902400	Neoplasm of uncertain behaviour of colon
	B902500	Neoplasm of uncertain behaviour of rectum
	BB5N.00	[M]Adenomatous and adenocarcinomatous polyps of colon
	BB5N.11	[M]Adenoma or or adenocarcinoma in polyposis coli
	BB5N100	[M]Adenocarcinoma in adenomatous polposis coli
	BB5Nz00	[M]Adenomatous or adenocarcinomatous polyps of the colon NOS
	ZV10017	[V]Personal history of malignant neoplasm of rectum

Enteropathic arthritis	14G1.00	H/O: rheumatoid arthritis
	F371200	Polyneuropathy in rheumatoid arthritis
	F396400	Myopathy due to rheumatoid arthritis
	G5y8.00	Rheumatoid myocarditis
	G5yA.00	Rheumatoid carditis
	H570.00	Rheumatoid lung
	M160.00	Psoriatic arthropathy
	M160.11	Psoriatic arthritis
	M160100	Distal interphalangeal psoriatic arthropathy
	M160z00	Psoriatic arthropathy NOS
	N013.00	Postdysenteric reactive arthropathy
	N013000	Postdysenteric reactive arthropathy of unspecified site
	N013100	Postdysenteric reactive arthropathy of the shoulder region
	N013200	Postdysenteric reactive arthropathy of the upper arm
	N013300	Postdysenteric reactive arthropathy of the forearm
	N013400	Postdysenteric reactive arthropathy of the hand
	N013500	Postdysenteric reactive arthropathy of the pelvis/thigh
	N013600	Postdysenteric reactive arthropathy of the lower leg
	N013700	Postdysenteric reactive arthropathy of the ankle and foot
	N013x00	Postdysenteric reactive arthropathy of multiple sites
	N013y00	Postdysenteric reactive arthropathy of other specified sites
	N013z00	Postdysenteric reactive arthropathy NOS
	N01w.00	Reactive arthropathy, unspecified
	N01w000	Reactive arthropathy of shoulder
	N01w100	Reactive arthropathy of sternoclavicular joint
	N01w200	Reactive arthropathy of acromioclavicular joint
	N01w300	Reactive arthropathy of elbow
	N01w400	Reactive arthropathy of distal radio-ulnar joint
	N01w500	Reactive arthropathy of wrist
	N01w600	Reactive arthropathy of MCP joint
	N01w700	Reactive arthropathy of PIP joint of finger
	N01w800	Reactive arthropathy of DIP joint of finger
	N01w900	Reactive arthropathy of hip
	N01wA00	Reactive arthropathy of sacro-iliac joint
	N01wB00	Reactive arthropathy of knee
	N01wC00	Reactive arthropathy of tibio-fibular joint
	N01wD00	Reactive arthropathy of ankle
	N01wE00	Reactive arthropathy of subtalar joint
	N01wF00	Reactive arthropathy of talonavicular joint
	N01wG00	Reactive arthropathy of other tarsal joint
	N01wH00	Reactive arthropathy of 1st MTP joint
	N01wJ00	Reactive arthropathy of lesser MTP joint
	N01wK00	Reactive arthropathy of IP joint of toe
	N031.00	Arthropathy assoc with non-infect gastrointestinal disorders
	N031000	Arthropathy in ulcerative colitis
	N031100	Arthropathy in Crohn's disease

N04..00	Rheumatoid arthritis and other inflammatory polyarthropathy
N040.00	Rheumatoid arthritis
N040000	Rheumatoid arthritis of cervical spine
N040100	Other rheumatoid arthritis of spine
N040200	Rheumatoid arthritis of shoulder
N040300	Rheumatoid arthritis of sternoclavicular joint
N040400	Rheumatoid arthritis of acromioclavicular joint
N040500	Rheumatoid arthritis of elbow
N040600	Rheumatoid arthritis of distal radio-ulnar joint
N040700	Rheumatoid arthritis of wrist
N040800	Rheumatoid arthritis of MCP joint
N040900	Rheumatoid arthritis of PIP joint of finger
N040A00	Rheumatoid arthritis of DIP joint of finger
N040B00	Rheumatoid arthritis of hip
N040C00	Rheumatoid arthritis of sacro-iliac joint
N040D00	Rheumatoid arthritis of knee
N040E00	Rheumatoid arthritis of tibio-fibular joint
N040F00	Rheumatoid arthritis of ankle
N040G00	Rheumatoid arthritis of subtalar joint
N040H00	Rheumatoid arthritis of talonavicular joint
N040J00	Rheumatoid arthritis of other tarsal joint
N040K00	Rheumatoid arthritis of 1st MTP joint
N040L00	Rheumatoid arthritis of lesser MTP joint
N040M00	Rheumatoid arthritis of IP joint of toe
N040N00	Rheumatoid vasculitis
N040P00	Seronegative rheumatoid arthritis
N040Q00	Rheumatoid bursitis
N040R00	Rheumatoid nodule
N040S00	Rheumatoid arthritis - multiple joint
N040T00	Flare of rheumatoid arthritis
N041.00	Felty's syndrome
N042.00	Other rheumatoid arthropathy + visceral/systemic involvement
N042000	Rheumatic carditis
N042100	Rheumatoid lung disease
N042200	Rheumatoid nodule
N042z00	Rheumatoid arthropathy + visceral/systemic involvement NOS
N043000	Juvenile rheumatoid arthropathy unspecified
N043100	Acute polyarticular juvenile rheumatoid arthritis
N043200	Pauciarticular juvenile rheumatoid arthritis
N043300	Monarticular juvenile rheumatoid arthritis
N043z00	Juvenile rheumatoid arthritis NOS
N045000	Juvenile ankylosing spondylitis
N045200	Juvenile arthritis in psoriasis
N045300	Juvenile arthritis in Crohn's disease
N045400	Juvenile arthritis in ulcerative colitis

	N045500	Juvenile rheumatoid arthritis
	N045600	Pauciarticular onset juvenile chronic arthritis
	N047.00	Seropositive erosive rheumatoid arthritis
	N04X.00	Seropositive rheumatoid arthritis, unspecified
	N04y000	Rheumatoid lung
	N04y011	Caplan's syndrome
	N04y012	Fibrosing alveolitis associated with rheumatoid arthritis
	N10..00	Inflammatory spondylopathies
	N100.00	Ankylosing spondylitis
	N100.11	Marie - Strumpell spondylitis
	N10z.00	Spondylitis NOS
	N11F.00	Axial spondyloarthritis
	N11y100	Enterobacterial spondylitis
	Nyu0300	[X]Other reactive arthropathies [X]Rheumatoid arthritis+involvement/other organs or systems
	Nyu1000	
	Nyu1100	[X]Other seropositive rheumatoid arthritis
	Nyu1200	[X]Other specified rheumatoid arthritis
	Nyu1300	[X]Other psoriatic arthropathies
	Nyu1400	[X]Other enteropathic arthropathies
	Nyu1G00	[X]Seropositive rheumatoid arthritis, unspecified
Depression	1465.00	H/O: depression
	1B17.00	Depressed
	1B17.11	C/O - feeling depressed
	1B1U.00	Symptoms of depression
	1B1U.11	Depressive symptoms
	1BO..00	Mood swings
	1BT..00	Depressed mood
	1BT..11	Low mood
	1BT..12	Sad mood
	2257.00	O/E - depressed
	225K.00	O/E - fearful mood
	9H90.00	Depression annual review
	9H91.00	Depression medication review
	9H92.00	Depression interim review
	9HA0.00	On depression register
	E112.00	Single major depressive episode
	E112.11	Agitated depression
	E112.12	Endogenous depression first episode
	E112.13	Endogenous depression first episode
	E112.14	Endogenous depression
	E112000	Single major depressive episode, unspecified
	E112100	Single major depressive episode, mild
	E112200	Single major depressive episode, moderate
	E112300	Single major depressive episode, severe, without psychosis
	E112400	Single major depressive episode, severe, with psychosis
	E112500	Single major depressive episode, partial or unspec remission

E112600	Single major depressive episode, in full remission
E112z00	Single major depressive episode NOS
E113.00	Recurrent major depressive episode
E113.11	Endogenous depression - recurrent
E113000	Recurrent major depressive episodes, unspecified
E113100	Recurrent major depressive episodes, mild
E113200	Recurrent major depressive episodes, moderate
E113300	Recurrent major depressive episodes, severe, no psychosis
E113400	Recurrent major depressive episodes, severe, with psychosis
E113500	Recurrent major depressive episodes, partial/unspecified remission
E113600	Recurrent major depressive episodes, in full remission
E113700	Recurrent depression
E113z00	Recurrent major depressive episode NOS
E11y200	Atypical depressive disorder
E11z100	Rebound mood swings
E11z200	Masked depression
E135.00	Agitated depression
E200300	Anxiety with depression
E204.00	Neurotic depression reactive type
E290.00	Brief depressive reaction
E290z00	Brief depressive reaction NOS
E291.00	Prolonged depressive reaction
E292400	Adjustment reaction with anxious mood
E2B..00	Depressive disorder NEC
E2B1.00	Chronic depression
Eu32.00	[X]Depressive episode
Eu32.11	[X]Single episode of depressive reaction
Eu32.12	[X]Single episode of psychogenic depression
Eu32.13	[X]Single episode of reactive depression
Eu32000	[X]Mild depressive episode
Eu32100	[X]Moderate depressive episode
Eu32200	[X]Severe depressive episode without psychotic symptoms
Eu32211	[X]Single episode agitated depression w/out psychotic symptoms
Eu32212	[X]Single episode major depression w/out psychotic symptoms
Eu32213	[X]Single episode vital depression w/out psychotic symptoms
Eu32300	[X]Severe depressive episode with psychotic symptoms
Eu32311	[X]Single episode of major depression and psychotic symptoms
Eu32312	[X]Single episode of psychogenic depressive psychosis
Eu32313	[X]Single episode of psychotic depression
Eu32314	[X]Single episode of reactive depressive psychosis
Eu32400	[X]Mild depression
Eu32500	[X]Major depression, mild
Eu32600	[X]Major depression, moderately severe
Eu32700	[X]Major depression, severe without psychotic symptoms

	Eu32800	[X]Major depression, severe with psychotic symptoms
	Eu32900	[X]Single major depr ep, severe with psych, psych in remiss
	Eu32A00	[X]Recurr major depr ep, severe with psych, psych in remiss
	Eu32y00	[X]Other depressive episodes
	Eu32y12	[X]Single episode of masked depression NOS
	Eu32z00	[X]Depressive episode, unspecified
	Eu32z11	[X]Depression NOS
	Eu32z12	[X]Depressive disorder NOS
	Eu32z13	[X]Prolonged single episode of reactive depression
	Eu32z14	[X] Reactive depression NOS
	Eu33.00	[X]Recurrent depressive disorder
	Eu33.11	[X]Recurrent episodes of depressive reaction
	Eu33.12	[X]Recurrent episodes of psychogenic depression
	Eu33.13	[X]Recurrent episodes of reactive depression
	Eu33000	[X]Recurrent depressive disorder, current episode mild
	Eu33100	[X]Recurrent depressive disorder, current episode moderate
	Eu33200	[X]Recurr depress disorder cur epi severe without psych sympt
	Eu33211	[X]Endogenous depression without psychotic symptoms
	Eu33212	[X]Major depression, recurrent without psychotic symptoms
	Eu33214	[X]Vital depression, recurrent without psychotic symptoms
	Eu33300	[X]Recurrent depress disorder cur epi severe with psych symp
	Eu33311	[X]Endogenous depression with psychotic symptoms
	Eu33313	[X]Recurr severe episodes/major depression+psychotic symptom
	Eu33314	[X]Recurr severe episodes/psychogenic depressive psychosis
	Eu33315	[X]Recurrent severe episodes of psychotic depression
	Eu33316	[X]Recurrent severe episodes/reactive depressive psychosis
	Eu33400	[X]Recurrent depressive disorder, currently in remission
	Eu33y00	[X]Other recurrent depressive disorders
	Eu33z00	[X]Recurrent depressive disorder, unspecified
	Eu33z11	[X]Monopolar depression NOS
	Eu34100	[X]Dysthymia
	Eu34111	[X]Depressive neurosis
	Eu34113	[X]Neurotic depression
	Eu34114	[X]Persistant anxiety depression
	Eu3y111	[X]Recurrent brief depressive episodes
	Eu41200	[X]Mixed anxiety and depressive disorder
	Eu41211	[X]Mild anxiety depression
Asthma	14B4.00	H/O: asthma
	173A.00	Exercise induced asthma
	173c.00	Occupational asthma
	173d.00	Work aggravated asthma
	178..00	Asthma trigger
	1780.00	Aspirin induced asthma
	1781.00	Asthma trigger - pollen
	1782.00	Asthma trigger - tobacco smoke

1783.00	Asthma trigger - warm air
1784.00	Asthma trigger - emotion
1785.00	Asthma trigger - damp
1786.00	Asthma trigger - animals
1787.00	Asthma trigger - seasonal
1788.00	Asthma trigger - cold air
1789.00	Asthma trigger - respiratory infection
178A.00	Asthma trigger - airborne dust
178B.00	Asthma trigger - exercise
1O2..00	Asthma confirmed
663e.00	Asthma restricts exercise
663e000	Asthma sometimes restricts exercise
663e100	Asthma severely restricts exercise
663f.00	Asthma never restricts exercise
663h.00	Asthma - currently dormant
663j.00	Asthma - currently active
663m.00	Asthma accident and emergency attendance since last visit
663N.00	Asthma disturbing sleep
663n.00	Asthma treatment compliance satisfactory
663N000	Asthma causing night waking
663N100	Asthma disturbs sleep weekly
663N200	Asthma disturbs sleep frequently
663O.00	Asthma not disturbing sleep
663O000	Asthma never disturbs sleep
663P.00	Asthma limiting activities
663p.00	Asthma treatment compliance unsatisfactory
663P000	Asthma limits activities 1 to 2 times per month
663P100	Asthma limits activities 1 to 2 times per week
663P200	Asthma limits activities most days
663q.00	Asthma daytime symptoms
663Q.00	Asthma not limiting activities
663r.00	Asthma causes night symptoms 1 to 2 times per month
663s.00	Asthma never causes daytime symptoms
663t.00	Asthma causes daytime symptoms 1 to 2 times per month
663u.00	Asthma causes daytime symptoms 1 to 2 times per week
663U.00	Asthma management plan given
663v.00	Asthma causes daytime symptoms most days
663V.00	Asthma severity
663V000	Occasional asthma
663V100	Mild asthma
663V200	Moderate asthma
663V300	Severe asthma
663w.00	Asthma limits walking up hills or stairs
663W.00	Asthma prophylactic medication used
663x.00	Asthma limits walking on the flat
66YC.00	Absent from work or school due to asthma

	66YJ.00	Asthma annual review
	66YK.00	Asthma follow-up
	66YP.00	Asthma night-time symptoms
	66Yq.00	Asthma causes night time symptoms 1 to 2 times per week
	66Yr.00	Asthma causes symptoms most nights
	66Ys.00	Asthma never causes night symptoms
	66Yu.00	Number days absent from school due to asthma in past 6 month
	H33..00	Asthma
	H33..11	Bronchial asthma
	H330.00	Extrinsic (atopic) asthma
	H330.11	Allergic asthma
	H330.12	Childhood asthma
	H330.13	Hay fever with asthma
	H330.14	Pollen asthma
	H330000	Extrinsic asthma without status asthmaticus
	H330011	Hay fever with asthma
	H330100	Extrinsic asthma with status asthmaticus
	H330111	Extrinsic asthma with asthma attack
	H330z00	Extrinsic asthma NOS
	H331.00	Intrinsic asthma
	H331.11	Late onset asthma
	H331000	Intrinsic asthma without status asthmaticus
	H331100	Intrinsic asthma with status asthmaticus
	H331111	Intrinsic asthma with asthma attack
	H331z00	Intrinsic asthma NOS
	H332.00	Mixed asthma
	H333.00	Acute exacerbation of asthma
	H334.00	Brittle asthma
	H335.00	Chronic asthma with fixed airflow obstruction
	H33z.00	Asthma unspecified
	H33z011	Severe asthma attack
	H33z100	Asthma attack
	H33z111	Asthma attack NOS
	H33z200	Late-onset asthma
	H33zz00	Asthma NOS
	H33zz11	Exercise induced asthma
	H33zz12	Allergic asthma NEC
	H35y700	Wood asthma
	H47y000	Detergent asthma
Dyspepsia	195..00	Indigestion symptoms
	1954.00	Indigestion
	1955.00	Heartburn
	1955.11	Heartburn symptom
	1958.00	Undiagnosed dyspepsia
	195Z.00	Indigestion symptom NOS
	8HI0.00	Referral to dyspepsia specialist nurse

	9NNK.00	Under care of dyspepsia specialist nurse
	E264400	Psychogenic dyspepsia
	Eu45318	[X]Psychogenic dyspepsia
	J101111	Acid reflux
	J10y412	Gastro-oesophageal reflux
	J10y413	Acid reflux
	J16y200	Hyperchlorhydria
	J16y211	Hyperacidity
	J16y400	Dyspepsia
	J16y411	Flatulent dyspepsia
	J16y412	Indigestion NOS
	J16yA00	Non-ulcer dyspepsia
	J16yA11	Functional dyspepsia
	R071.00	[D]Heartburn
	R071z00	[D]Heartburn NOS
Change in bowel habit	19EA.00	Change in bowel habit
	19EA.11	Altered bowel habit
	19EE.00	Increased frequency of defaecation
	J521200	IBS characterised by alternating bowel habit
	R078.00	[D]Change in bowel habit
Constipation	19EA.00	Change in bowel habit
	19C..00	Constipation
	19C..11	Constipation symptom
	19C..12	Costive symptom
	19C2.00	Constipated
	19CZ.00	Constipation NOS
	E264500	Psychogenic constipation
	J520.00	Constipation - functional
	J520000	Acute constipation
	J520100	Chronic constipation with overflow
	J520200	Chronic constipation without overflow
	J520400	Chronic constipation
	J520y00	Other specified constipation
	J520z00	Constipation NOS
	J521100	Irritable bowel syndrome characterised by constipation
Diarrhoea	J52y100	Difficulty in ability defaecat
	19E3.00	Incontinent of faeces
	19E3.11	Incontinent of faeces symptom
	19EE.00	Increased frequency of defaecation
	19EF.00	Urgent desire for stool
	19F..00	Diarrhoea symptoms
	19F..11	Diarrhoea
	19F..12	Loose stools
	19F2.00	Diarrhoea
	19F3.00	Spurious (overflow) diarrhoea
	19F5.00	Time since last episode of diarrhoea

	19FZ.00	Diarrhoea symptom NOS
	19FZ.11	Diarrhoea & vomiting, symptom
	19G..00	Diarrhoea and vomiting
	E264300	Psychogenic diarrhoea
	E264311	Spurious diarrhoea
	Eu45317	[X]Psychogenic diarrhoea
	J4...13	Noninfective diarrhoea
	J43z.11	Chronic diarrhoea
	J4z..11	Presumed noninfectious diarrhoea
	J4zz.11	Diarrhoea - presumed non-infectious
	J521000	Irritable bowel syndrome with diarrhoea
	J525.00	Functional diarrhoea
	J528.00	Intestinal hurry
	R077100	[D] Stools loose
Loperamide	52020978	Loperamide 2mg capsules
	67054979	Loperamide 2.5mg/5ml oral solution
	73286978	Loperamide 2mg capsules
	79611979	Loperamide 4mg/5ml oral solution
	79615979	Loperamide 25mg/5ml oral solution
	81123998	Loperamide 25mg/5ml oral suspension
	83351978	Loperamide 2mg capsules
	83494978	Loperamide 2mg / simeticone 125mg tablets
	99505979	Loperamide 2mg capsules
	46952978	Loperamide hydrochloride 2mg oro-dispersible tablets
	57011978	Loperamide 2mg orodispersible tablets sugar free
	59423978	Loperamide 2mg orodispersible tablets sugar free
	59424978	Loperamide 2mg orodispersible tablets sugar free
	67059979	Loperamide 1mg/5ml oral solution sugar free
	73285978	Loperamide 2mg capsules
	82016998	Loperamide 2mg capsules
	82017998	Loperamide 2mg capsules
	83183998	Loperamide 2mg capsules
	86504998	Loperamide 2mg capsules
	88039998	Loperamide 2mg / Simeticone 125mg tablets
	88040998	Loperamide 2mg / Simeticone 125mg tablets
	88041998	Loperamide 2mg / Simeticone 125mg tablets
	88353998	Loperamide 2mg / Simeticone 125mg tablets
	88544998	Loperamide 2mg / Simeticone 125mg tablets
	88731998	Loperamide hydrochloride 2mg capsules
	89502997	Loperamide 2mg capsules
	89570997	Loperamide 1mg/5ml oral solution sugar free
	89570998	Loperamide 2mg capsules
	89987998	Loperamide 2mg tablets
	90575998	Loperamide 2mg capsules
	90715998	Loperamide 2mg tablets
	91167998	Loperamide hydrochloride 2mg capsules

	91224998	Loperamide and oral rehydration salts capsule and sachets
	92039998	Loperamide 2mg capsules
	92474998	Loperamide 2mg orodispersible tablets sugar free
	92475998	Loperamide hydrochloride 2mg oro-dispersible tablets
	92476998	Loperamide 2mg orodispersible tablets sugar free
	92746998	Loperamide hydrochloride 2mg capsules
	95996998	Loperamide 1mg/5ml oral solution sugar free
	95997997	Loperamide 2mg tablets
	95997998	Loperamide 2mg capsules
	96814998	Loperamide 2mg capsules
	96854990	Loperamide 2mg capsules
	98036990	Loperamide 2mg capsules
	98037990	Loperamide 2mg capsules
	98963990	Loperamide 2mg capsules
	98964990	Loperamide 2mg capsules
	99537997	Loperamide hydrochloride 1mg/5ml sugar free liquid
	99537998	Loperamide 2mg capsules
Kaolin	89502998	Kaolin 400mg / calcium carbonate 75mg tablets
	89503998	Kaolin 400mg / calcium carbonate 75mg tablets
	96041998	Kaolin with pectin and sodium citrate mixture
	96042998	Kaolin with pectin and sodium citrate mixture
	97636992	Kaolin light 6.18 gm mix
	97637992	Kaolin comp mix
	97907998	Kaolin mixture
	98047998	Kaolin mixture
	99742990	Kaolin mixture
	90230998	Kaolin 500mg/5ml / Calcium carbonate 250mg/5ml oral suspension sugar free
	90231998	Kaolin 500mg/5ml / Calcium carbonate 250mg/5ml oral suspension sugar free
	93122998	Morphine hcl light kaolin, belladonna and aluminium hydroxide chewable tablet
	93282998	Morphine hcl and light kaolin and calcium carbonate tablet
	93291998	Morphine hcl and kaolin and belladonna tablet
	94104998	Kaolin light 1.5g/5ml / codeine 5mg/5ml oral suspension sugar free
	98049998	Kaolin and Morphine mixture
	99518998	Codeine phosphate & kaolin 10mg+3g/10ml mixture
	99542990	Kaolin and Morphine mixture
Racecadotril	53801979	Racecadotril 30mg granules sachets
	53802979	Racecadotril 30mg granules sachets
	53803979	Racecadotril 10mg granules sachets
	53804979	Racecadotril 10mg granules sachets
	53805979	Racecadotril 100mg capsules
	53806979	Racecadotril 100mg capsules
Bloating	19A..00	Abdominal distension symptom
	19A2.00	Abdomen feels bloated
	19A3.00	Abdomen feels distended
	19A4.00	Abdomen feels swollen

	19AZ.00	Abd. distension symptom NOS
	19B..00	Flatulence/wind
	19B..12	Bloating symptom
	19B..14	Flatulence symptom
	19B..15	Wind symptom
	19B2.00	Excessive flatulence
	19B5.00	Excessive flatus
	19BZ.00	Wind NOS
	Eu45320	[X]Psychogenic flatulence
	R073.00	[D]Flatulence, eructation and gas pain
	R073000	[D]Flatulence
	R073200	[D]Gas pain (abdominal)
	R073300	[D]Abdominal distension, gaseous
	R073400	[D]Bloating
	R073500	[D]Tympanites (abdominal)
	R073z00	[D]Flatulence, eructation and gas pain NOS
	R073z11	[D]Wind
Abdominal pain	196..11	Abdominal pain type
	1962.00	Colicky abdominal pain
	1963.00	Non-colicky abdominal pain
	1967.00	Abdominal migraine - symptom
	1968.00	Abdominal discomfort
	1969.00	Abdominal pain
	1969000	Abdominal wall pain
	197..11	Flank pain
	197..12	Iliac fossa pain
	197..13	Site of abdominal pain
	197..14	Subcostal pain
	1971.00	Central abdominal pain
	1972.00	Epigastric pain
	1973.00	Left subcostal pain
	1974.00	Right subcostal pain
	1975.00	Left flank pain
	1976.00	Right flank pain
	1977.00	Right iliac fossa pain
	1978.00	Left iliac fossa pain
	1979.00	Suprapubic pain
	197A.00	Generalised abdominal pain
	197A.11	General abdominal pain-symptom
	197B.00	Upper abdominal pain
	197C.00	Lower abdominal pain
	197D.00	Right upper quadrant pain
	25C..00	O/E - abdo. pain on palpation
	25C..11	O/E - epigastric pain on palp.
	25C..12	O/E - iliac pain on palpation
	25C..14	O/E - umbilical pain on palp.

	25C..15	O/E - abdomen tender
	25C2.00	O/E - abd.pain-R.hypochondrium
	25C3.00	O/E - abd. pain - epigastrium
	25C4.00	O/E - abd.pain-L.hypochondrium
	25C5.00	O/E - abd. pain - R.lumbar
	25C6.00	O/E - abd. pain - umbilical
	25C7.00	O/E - abd. pain - L.lumbar
	25C8.00	O/E - abd. pain - R.ilic
	25C9.00	O/E - abd. pain - hypogastrium
	25CA.00	O/E - abd. pain - L.ilic
	25CZ.00	O/E -abd.pain on palpation NOS
	F262200	Abdominal migraine
	R073200	[D]Gas pain (abdominal)
	R090.00	[D]Abdominal pain
	R090000	[D]Abdominal tenderness
	R090100	[D]Abdominal colic
	R090400	[D]Abdominal cramps
	R090500	[D]Epigastric pain
	R090600	[D]Umbilical pain
	R090700	[D]Hypochondrial pain
	R090800	[D]Suprapubic pain
	R090900	[D]Pain in right iliac fossa
	R090A00	[D]Pain in left iliac fossa
	R090C00	[D]Loin pain
	R090D00	[D]Abdominal migraine
	R090E00	[D]Recurrent acute abdominal pain
	R090F00	[D]Acute abdomen
	R090G11	[D] Pelvic pain
	R090H00	[D]Upper abdominal pain
	R090J00	[D]Right upper quadrant pain
	R090K00	[D]Left upper quadrant pain
	R090L00	[D]Left lower quadrant pain
	R090M00	[D]Right lower quadrant pain
	R090N00	[D]Nonspecific abdominal pain
	R090P00	[D]Functional abdominal pain syndrome
	R090y00	[D]Other specified abdominal pain
	R090z00	[D]Abdominal pain NOS
	Ryu1000	[X]Pain localized to other parts of lower abdomen
	Ryu1100	[X]Other and unspecified abdominal pain
FC test	47J..00	Faecal calprotectin content
	47J0.00	Faecal calprotectin test indeterminate
	47J1.00	Faecal calprotectin test invalid
	47J2.00	Faecal calprotectin test positive
	47J3.00	Faecal calprotectin test negative
Colonoscopy	4Q3J.00	Calprotectin level
	3614200	Proctoscopy normal

	3615200	Sigmoidoscopy abnormal
	3615300	Proctoscopy abnormal
	3617.00	Colonoscopy normal
	3618.00	Colonoscopy abnormal
	3619.00	Colonoscopy equivocal
	361A.00	Sigmoidoscopy normal
	7637.11	Diagnostic jejunoscopy
	7637z11	Jejunoscopy NEC
	764A.11	Diagnostic enteroscopy
	764A.12	Diagnostic ileoscopy
	771..00	Colon operations and sigmoidoscopy of rectum
	771E.11	Open colonoscopy
	771E000	Open colonoscopy
	771J.11	Diagnostic colonoscopy
	771J100	Check colonoscopy
	771J200	Limited colonoscopy
	771M100	Sigmoidoscopy NEC
	771Mz11	Fibreoptic sigmoidoscopy NEC
	771Qz11	Sigmoidoscopy NEC
	771y.00	Other specified operations on colon or rectal sigmoidoscopy
	771z.00	Colon operations or rectal sigmoidoscopy NOS
	772A.00	Diagnostic proctoscopy
	772A000	Diagnostic proctoscopy and biopsy of lesion of rectum
	772Ay00	Other specified diagnostic proctoscopy
	772Az00	Diagnostic proctoscopy NOS
	772Bz11	Proctoscopy NEC
	8H5J.00	Referral to colorectal surgeon
	8HS0.00	Refer for sigmoidoscopy
	8HS0000	Referral to community sigmoidoscopy service
	8HU1.00	Referral for colonoscopy
	8HVc.00	Private referral to colorectal surgeon
	8LJ..00	Colonoscopy planned
	9EV5.00	Colonoscopy report received
	ZL5GD00	Referral to colorectal surgeon
	ZV58711	[V]Sigmoidoscopy normal
	1001400169	Colonoscopy
	1001400170	Sigmoidoscopy
Mesalazine	53913979	Mesalazine 800mg gastro-resistant tablets
	88517998	Mesalazine 400mg gastro-resistant tablets
	93624997	Mesalazine 250mg modified release tablets
	99583998	Mesalazine 250mg modified-release tablet
	54552979	Mesalazine 400mg gastro-resistant tablets
	55164978	Mesalazine 4g modified-release granules sachets sugar free
	55165978	Mesalazine 4g modified-release granules sachets sugar free
	58800979	Mesalazine 1.2g gastro-resistant modified-release tablets
	60584979	Mesalazine 3g gastro-resistant modified-release granules sachets sugar free

60585979	Mesalazine 3g gastro-resistant modified-release granules sachets sugar free
64868979	Mesalazine 2g modified-release granules sachets sugar free
76424978	Mesalazine 1g modified-release tablets
79867978	Mesalazine 400mg gastro-resistant tablets
80928998	Mesalazine 3g gastro-resistant modified-release granules sachets sugar free
80929998	Mesalazine 3g gastro-resistant modified-release granules sachets sugar free
81193998	Mesalazine 1g modified-release tablets
81194998	Mesalazine 1g modified-release tablets
81683998	Mesalazine 1g suppositories
81689998	Mesalazine 500mg gastro-resistant tablets
81690998	Mesalazine 500mg gastro-resistant tablets
81772998	Mesalazine 800mg gastro-resistant tablets
81868998	Mesalazine 400mg gastro-resistant tablets
83503998	Mesalazine 1.5g gastro-resistant modified-release granules sachets sugar free
83504998	Mesalazine 1.5g gastro-resistant modified-release granules sachets sugar free
83743998	Mesalazine 2g modified-release granules sachets sugar free
83987998	Mesalazine 2g modified-release granules sachets sugar free
84059998	Mesalazine 1g gastro-resistant modified-release granules sachets sugar free
84209998	Mesalazine 800mg gastro-resistant tablets
84290998	Mesalazine 1.2g gastro-resistant modified-release tablets
84291998	Mesalazine 1.2g gastro-resistant modified-release tablets
85560998	Mesalazine 800mg gastro-resistant tablets
87761998	Mesalazine 400mg gastro-resistant tablets
87909998	Mesalazine 1g gastro-resistant modified-release granules sachets sugar free
87910998	Mesalazine 500mg gastro-resistant modified-release granules sachets sugar free
87911998	Mesalazine 500mg gastro-resistant modified-release granules sachets sugar free
89992997	Mesalazine 2g/59ml enema
89992998	Mesalazine 500mg suppositories
92346998	Mesalazine 1g/application foam enema
92347998	Mesalazine 400mg gastro-resistant tablets
92764997	Mesalazine 1g modified-release granules sachets sugar free
92764998	Mesalazine 500mg modified-release tablets
93623996	Mesalazine 250mg suppositories
93623997	Mesalazine 1g/application foam enema
93623998	Mesalazine 1g/100ml enema
93624996	Mesalazine 1g suppositories
93624998	Mesalazine 1g/100ml enema
93728992	Mesalazine 500mg modified-release tablets
94564992	Mesalazine 500mg modified-release tablets
95041990	Mesalazine 400mg gastro-resistant tablets
95888997	Mesalazine 250mg gastro-resistant tablets
95888998	Mesalazine 400mg gastro-resistant tablets
96608996	Mesalazine 2g/59ml enema

	96608997	Mesalazine 1g suppositories
	96608998	Mesalazine 500mg suppositories
	96659996	Mesalazine 1g/application foam enema
	96659997	Mesalazine 500mg suppositories
	96659998	Mesalazine 250mg suppositories
	96883990	Mesalazine 400mg gastro-resistant tablets
	96916992	Mesalazine 500mg modified-release tablets
	97381998	Mesalazine 400mg gastro-resistant tablets
	97764998	Mesalazine 250mg gastro-resistant tablets
	98001992	Mesalazine 250mg gastro-resistant tablets
	99486979	Mesalazine 400mg gastro-resistant tablets
	99487979	Mesalazine 400mg gastro-resistant tablets
	99488979	Mesalazine 400mg gastro-resistant tablets
	99490979	Mesalazine 1g modified-release granules sachets sugar free
	99492979	Mesalazine 1g suppositories
	99494979	Mesalazine 1g suppositories
	99495979	Mesalazine 500mg modified-release tablets
	99498979	Mesalazine 500mg modified-release tablets
		Mesalazine 1g gastro-resistant modified-release granules sachets sugar free
	99583996	
	99583997	Mesalazine 500mg modified-release tablets
Olsalazine	92400998	Olsalazine 500mg tablets
	92401998	Olsalazine 250mg capsules
	94437997	Olsalazine 500mg tablets
	94437998	Olsalazine 250mg capsules
	94438997	Olsalazine 500mg tablets
	94438998	Olsalazine 250mg capsules
Balsalazide	88489998	Balsalazide 750mg capsules
	88492998	Balsalazide 750mg capsules
Sulfasalazine	89598997	Sulfasalazine 3g/100ml retention enema
	89598998	Sulfasalazine 500mg suppositories
	89604997	Sulfasalazine 3g/100ml retention enema
	89604998	Sulfasalazine 500mg suppositories
	95256996	Sulfasalazine 3g/100ml enema
	95256997	Sulfasalazine 500mg suppositories
	97281996	Sulfasalazine 3g/100ml retention enema
	97281997	Sulfasalazine 500mg suppositories
Beclometasone		Beclometasone 5mg gastro-resistant modified-release tablets
	86941998	Beclometasone 5mg gastro-resistant modified-release tablets
	86942998	Beclometasone 5mg gastro-resistant modified-release tablets
Budesonide	50988978	Budesonide 9mg modified-release tablets
	50989978	Budesonide 9mg modified-release tablets
	60586979	Budesonide 9mg gastro-resistant granules sachets
	60587979	Budesonide 9mg gastro-resistant granules sachets
	80931998	Budesonide 9mg gastro-resistant granules sachets
	80932998	Budesonide 9mg gastro-resistant granules sachets
	84636998	Budesonide 2mg foam enema

Prednisolone	84637998	Budesonide 2mg foam enema
	89238998	Budesonide 2mg/100ml enema
	89239998	Budesonide 2mg/100ml enema
	94125992	Cortenema enema ml liq
	60097979	Prednisolone 20mg/application foam enema
	84741998	Prednisolone 40mg/100ml enema
	90310979	Prednisolone 20mg/application foam enema
	93706998	Prednisolone sodium phosphate 5mg suppositories
	94336997	Prednisolone sodium phosphate 5mg suppositories
	94336998	Prednisolone rectal ointment
	94451998	Prednisolone 20mg/application foam enema
	94452998	Prednisolone 20mg/application foam enema
	94468998	Prednisolone 20mg/100ml enema standard tube
	98370998	Prednisolone sodium phosphate 5mg suppositories
	98371998	Prednisolone 20mg/100ml enema standard tube
	99227997	Prednisolone 20mg/100ml enema standard tube
	99227998	Prednisolone 20mg/100ml enema standard tube
	99388979	Prednisolone 20mg/100ml enema standard tube
	99389979	Prednisolone 20mg/100ml enema standard tube
	89284997	Cinchocaine 1mg / Prednisolone hexanoate 1.3mg suppositories

A4 EXPLORATION OF THE THIN DATASET AND MANAGEMENT DECISIONS OF IRREGULAR CASES

Table A2 summarises illogical or inaccurate cases with an IBD record. 561 patients had an IBD diagnosis date prior or equal to their year of birth. 8,739 patients had an IBD Read code date prior to 1985, the approximate date of the introduction of the Read code system.²¹³ 24 patients had an IBD diagnosis recorded based on the prescription of IBD specific medication dating back to before 1945, while the medications to resolve IBD symptoms were first discovered in the 1940s.²¹⁵

Table A2 Frequencies of IBD records identified as illogical or inaccurate

Characteristic of IBD diagnosis date	Number of cases (% of 86,694 people with IBD diagnosis)
IBD diagnosis date < YOB	527 (0.6)
IBD diagnosis date = 1899	1 (0.0)
IBD diagnosis date = 1900	502 (0.6)
IBD diagnosis date = 1901	16 (0.0)
IBD diagnosis date = 1908	4 (0.0)
IBD diagnosis date = YOB	34 (0.0)

IBD diagnosis by Read code < 1985	8739 (10.0)
IBD diagnosis by drug code < 1945	24 (0.0)

YOB year of birth

Patient records in the THIN datasets are a combination of 1) contemporary records made by the THIN contributing practice at or after the patient's registration and 2) events that occurred before the patient registered with that practice.

The irregular dates are events that occurred before the patient registered at the THIN practice, which were corrupted when data were transferred from other clinical systems to the Vision software. They are true events but the date is imprecise. Because these events can be treated as true events the data are informative for the analysis of incidence and prevalence.

Because the date is not reliable we need to clarify whether knowing the true date might change the estimates of incidence and/or prevalence, i.e. whether using the imprecise dates might lead to wrong estimates.

The dates in question concern diagnosis dates before the patient registered at the THIN practice and therefore occurred before the start of the observation period. Incidence rates only take into account new cases. Any patients with a previous diagnosis are excluded because they are no longer at risk. For prevalence, any past diagnoses will be considered from the start of the observation period. The true date is irrelevant for the calculation. Therefore, I included all patients with an irregular IBD diagnosis date into the study of prevalence and incidence.

Appendix B Test accuracy of FC testing for IBD

B1 SEARCH STRATEGY

Ovid MEDLINE(R) 1946 to May Week 3 2017 (searched on 31/05/2017)
472 references exported

1. exp Inflammatory Bowel Diseases/di [Diagnosis]	10670
2. exp Irritable Bowel Syndrome/di [Diagnosis]	1303
3. Crohn* disease.tw.	34506
4. Ulcerative colitis.tw.	29458
5. Inflammatory bowel disease.tw.	27920
6. Irritable bowel syndrome.tw.	9092
7. (IBS or IBD).tw.	20353
8. (chronic diarrhoea or chronic diarrhea).tw.	3608
9. (abdominal pain or abdominal discomfort).tw.	42241
10. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9	124837
11. Calprotectin.tw.	1463
12. exp Leukocyte L1 Antigen Complex/an [Analysis]	606
13. 11 or 12	1594
14. 10 and 13	752
15. Limit 14 to ed=20120901-20170531	472

Ovid MEDLINE(R) Daily Update May 30, 2017, Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations May 30, 2017, Ovid MEDLINE(R) Epub Ahead of Print May 30, 2017 (searched on 31/05/2017)
240 references exported

1. exp Inflammatory Bowel Diseases/di [Diagnosis]	15
2. exp Irritable Bowel Syndrome/di [Diagnosis]	3
3. Crohn* disease.tw.	4163
4. Ulcerative colitis.tw.	3292
5. Inflammatory bowel disease.tw.	4537
6. Irritable bowel syndrome.tw.	1568
7. (IBS or IBD).tw.	4167
8. (chronic diarrhoea or chronic diarrhea).tw.	332
9. (abdominal pain or abdominal discomfort).tw.	7253
10. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9	17922
11. Calprotectin.tw.	435
12. exp Leukocyte L1 Antigen Complex/an [Analysis]	3
13. 11 or 12	435
14. 10 and 13	240

Embase 1974 to 2017 May 30 (searched on 31/05/2017)
1995 references exported

1. crohn* disease.tw.	58919
2. ulcerative colitis.tw.	47316
3. inflammatory bowel disease*.tw.	56906
4. irritable bowel syndrome*.tw.	16123
5. (IBD or IBS).tw.	45994
6. (chronic diarrhoea or chronic diarrhea).tw.	5437
7. (abdominal pain or abdominal discomfort).tw.	77520
8. 1 or 2 or 3 or 4 or 5 or 6 or 7	213302
9. calprotectin.tw.	3698
10. exp calgranulin/ec [Endogenous Compound]	2298
11. 9 or 10	4798
12. 8 and 11	2731
13. exp Crohn disease/di [Diagnosis]	9738
14. exp ulcerative colitis/di [Diagnosis]	6189
15. 13 or 14	13053
16. 11 and 15	347
17. 12 or 16	2747
18. (fecal or faecal).tw.	82659
19. 17 and 18	2159
20. limit 19 to dd=20120901-20170531	1487
21. limit 19 to em=201235-201722	1975
22. 20 or 21	1995

Cochrane library (search on 31/05/2017)

calprotectin and (inflammatory bowel disease* or irritable bowel syndrome or crohn* disease or ulcerative colitis) - 158 hits

Other reviews: 6

Trials: 147

Technology assessments: 3

Economic evaluations: 2 (not of interest)

WoS: Science Citation Index and Conference Proceedings Citation Index 2012-2017 (search on 31/05/2017)

TS=(calprotectin and (inflammatory bowel disease* or irritable bowel syndrome or crohn* disease or ulcerative colitis)) – 1,063 hits

B2 BETWEEN STUDY VARIANCE OF LOGIT SENSITIVITY AND LOGIT SPECIFICITY OF EACH RANDOM EFFECTS META-ANALYSIS (FOR TESTS AND CLINICAL QUESTIONS)

There was some statistical evidence (Chi-square = 14.145, 9df, $p=0.1173$) that the assumption of equal variances for tests is reasonable at the significance level of 0.05. However, the number of studies per test was small, and examination of the variances of the tests revealed that there is a substantial difference in the between-study variance of logit sensitivity and logit specificity of each test (Table B1). The model did not converge when unequal variances were assumed between test types.

For clinical questions there was some statistical evidence (Chi-square = 13.328, 6df, $P=0.03812$) that the assumption of equal variance may not be reasonable for the 50 μ g/g cut-off, but reasonable for the 100 μ g/g cut-off (Chi-square = 2.3185, 6df, $P=0.8882$) with a much smaller number of studies. However, the models with non-equal variances did not converge, so results should be interpreted with caution.

Table B1 Between study variance of logit sensitivity and logit specificity for tests and clinical questions

Test name	Between study variance (logitSe)	Between study variance (logitSp)	Correlation between logitSe and logitSp
PhiCal	1.301	1.202	-0.94
EK-CAL	0.506	1.265	-0.47
Quantum-Blue	2.2012	0.1378	1.00
ELiA	0.9620	0.6593	0.132
Clinical questions			
IBD vs IBS	1.4131	0.4883	-0.81
IBD vs non-IBD	1.1198	0.4731	-0.81
Organic vs non-organic disease	0.1546	0.7890	0.48

IBD inflammatory bowel disease, IBS irritable bowel syndrome, logitSe logit sensitivity, logitSp logit specificity

B3 R CODE FOR MY EXPLORATORY 25,000 META-ANALYSES AND 2D DENSITY PLOT

```
#reading in studies
study.names <- c("study1", "study2", ... ,"study 28")
n.study <- 28

#-----

# number of 'bootstraps'
boots <- 25001
#taking studies at random and if run again will take same random studies and produce the same
results
set.seed(1234)
sensitivity <- vector(length=boots)
specificity <- vector(length=boots)

#-----

#start of loop
for(j in 1:boots){
  data.meta <- NULL
  for(i in 1:n.study){
    pick.study <- df[df$'Study ID' == study.names[i], ]
    number.rows <- dim(pick.study)[1]
    pick.it <- sample(1:number.rows, 1, replace = FALSE)
    #print(pick.study)
    #print(number.rows)
    #print(pick.study[pick.it, ])
    data.meta <- rbind(data.meta, pick.study[pick.it, ])
  } # end i

#-----

# start of meta-analysis
#adding variables (FN= false negatives, TN=true negatives, TP=true positives, FP=false positives from
your studies recorded in dataframe df)
data.meta$n1 <- data.meta$TP+data.meta$FN
data.meta$n0 <- data.meta$FP+data.meta$TN
data.meta$true1 <- data.meta$TP
data.meta$true0 <- data.meta$TN
data.meta$study <- 1:28

#-----
```

```

#reshape data
Y= reshape(data.meta, direction = "long", varying = list( c("n1" , "n0" ) ,
                                                         c( "true1","true0" ) ) , timevar = "sens" , times = c(1,0) ,
           v.names = c("n","true" ) )
Y= Y[order(Y$id),]
Y$spec<- 1-Y$sens

#-----

# Bivariate random-effects meta-analysis104 of 28 studies randomly picked
(MA_Y = glmer( formula = cbind( true , n - true ) ~ 0 + sens + spec + (0+sens + spec|study),
              data = Y , family = binomial))
#test accuracy measures and CI on the raw scale
(ma_Y = summary(MA_Y))
labels( ma_Y)
ma_Y$coeff
(Isens = ma_Y$coeff[1,1])
(Ispec = ma_Y$coeff[2,1])

sensitivity[j] <- plogis(Isens)
specificity[j] <- plogis(Ispec)
#sensitivity[j] <- plogis( Sens )
#specificity[j] <- plogis( Spec )

#end of MA

#-----

print(data.meta)
} # end j

#-----

#preparing data for plotting: create columns of sensitivity and 1-specificity (false positive rate)
FPR <- 1-specificity
all.pairs <- cbind(sensitivity, FPR)

#-----

#2D density plot
#create a dataframe
all.pairs.df <- data.frame(all.pairs)
library(ggplot2)
commonTheme = list(labs(color="Density",fill="Density",

```

```

    x="1-specificity",
    y="sensitivity"),
  theme_bw(),
  theme(legend.position=c(1,1),
        legend.justification=c(1,1)))

ggplot(all.pairs.df,aes(1-specificity, sensitivity))+
  stat_density2d(aes(fill=..level..,alpha=..level..),geom='polygon',colour='black')+
  scale_fill_continuous(low="green",high="red")+
  guides(alpha="none")+
  commonTheme+
  coord_fixed(1,xlim=c(0.2125,0.275), ylim=c(0.8625,0.9375))

#-----

#scatter in ROC space with line of no discrimination

ggplot(all.pairs.df, aes(`1-specificity`,sensitivity))+
  coord_fixed(1,xlim=c(0,1), ylim=c(0,1))+
  geom_point(shape=1)+
  geom_abline(intercept = 0, slope = 1)

```

B4 EXCLUDED STUDIES WITH REASON

Table B2 Full texts excluded from review

Reference of full text	Reason for exclusion
1. Anonymous. What is the faecal calprotectin test? Drug & Therapeutics Bulletin. 2014;52(9):102-4.	Review
2. Bai W, Boswell T. Clinical utility and outcome analysis of faecal calprotectin in Hawkes Bay District Health Board. New Zealand Medical Journal. 2016;129(1433):69-73.	No 2x2 data, 46/85 patients with FC test for differential diagnosis (IBD versus non-IBD) had no final diagnosis reported, FP only partially reported for patients with IBS diagnosis, not for non-IBD diagnoses, FN for IBD not reported
3. Bar-Gil Shitrit A, Koslowsky B, Livovsky DM, Shitrit D, Paz K, Adar T, Adler SN, Goldin E. A prospective study of fecal calprotectin and lactoferrin as predictors of small bowel Crohn's disease in patients undergoing capsule endoscopy. Scandinavian Journal of Gastroenterology. 2017;52(3):328-33.	Small bowel only, patient spectrum - all patients included had previous negative colonoscopy

<p>4. Burri E, Manz M, Rothen C, Rossi L, Beglinger C, Lehmann FS. Monoclonal antibody testing for fecal calprotectin is superior to polyclonal testing of fecal calprotectin and lactoferrin to identify organic intestinal disease in patients with abdominal discomfort. <i>Clinica Chimica Acta</i>. 2013;416:41-7.</p>	<p>Overlap with previous SR</p>
<p>5. Burri E, Manz M, Schroeder P, Froehlich F, Rossi L, Beglinger C, Lehmann FS. Diagnostic yield of endoscopy in patients with abdominal complaints: incremental value of faecal calprotectin on guidelines of appropriateness. <i>BMC Gastroenterology</i>. 2014;14:57.</p>	<p>2x2 data are for risk of significant endoscopy finding by EPAGE alone versus EPAGE plus FC (not FC versus actual endoscopy finding)</p>
<p>6. Chang MH, Chou JW, Chen SM, Tsai MC, Sun YS, Lin CC, Lin CP. Faecal calprotectin as a novel biomarker for differentiating between inflammatory bowel disease and irritable bowel syndrome. <i>Molecular Medicine Reports</i>. 2014;10(1):522-6.</p>	<p>Patients with confirmed IBD/ IBS and healthy controls included, no 2x2 data, sensitivity and specificity and predictive values reported for 'optimal' cut off from ROC for IBD patients only</p>
<p>7. Chapman TP, Chen LY, Leaver L. Investigating young adults with chronic diarrhoea in primary care. <i>BMJ</i>. 2015;350:h573.</p>	<p>Case study, report on best clinical practice</p>
<p>8. Egea Valenzuela J, Pereniguez Lopez A, Perez Fernandez V, Alberca de Las Parras F, Carballo Alvarez F. Fecal calprotectin and C-reactive protein are associated with positive findings in capsule endoscopy in suspected small bowel Crohn's disease. <i>Revista Espanola de Enfermedades Digestivas</i>. 2016;108(7):394-400.</p>	<p>Small bowel only, spectrum bias - all patients included had previous negative colonoscopy</p>
<p>9. Egea-Valenzuela J, Alberca-de-Las-Parras F, Carballo-Alvarez F. Fecal calprotectin as a biomarker of inflammatory lesions of the small bowel seen by videocapsule endoscopy. <i>Revista Espanola de Enfermedades Digestivas</i>. 2015;107(4):211-4.</p>	<p>Small bowel only, spectrum bias - all patients included had previous negative colonoscopy</p>
<p>10. Elias SG, Kok L, de Wit NJ, Witteman BJ, Goedhard JG, Romberg-Camps MJ, Muris JW, Moons KG. Is there an added value of faecal calprotectin and haemoglobin in the diagnostic work-up for primary care patients suspected of significant colorectal disease? A cross-sectional diagnostic study. <i>BMC Medicine</i>. 2016;14(1):141.</p>	<p>Risk prediction model modelling the incremental diagnostic accuracy of FC test in addition to physical examination predictors</p>

<p>11. Emmanuel A, Landis D, Peucker M, Hungin AP. Faecal biomarker patterns in patients with symptoms of irritable bowel syndrome. <i>Frontline Gastroenterology</i>. 2016;7(4):275-82.</p>	<p>FC for characterisation of demographics and IBS subtypes in patients with known IBS</p>
<p>12. Garcia F, Martinez C, Juliao H, Bautista-Molano W, Valle-Onate R, Rueda JC, Romero-Sanchez C. Autoantibodies and fecal calprotectin levels in a group of Colombian patients with inflammatory bowel disease. <i>Gazzetta Medica Italiana Archivio per le Scienze Mediche</i>. 2017;176(3):132-41.</p>	<p>IBD patients only, IBS excluded, not test accuracy but characterisation of severity of IBD patients</p>
<p>13. Hale MF, Drew K, McAlindon ME, Sidhu R. The diagnostic accuracy of faecal calprotectin and small bowel capsule endoscopy and their correlation in suspected isolated small bowel Crohn's disease. <i>European Journal of Gastroenterology & Hepatology</i>. 2016;28(10):1145-50.</p>	<p>Small bowel only, unclear whether included patients had normal colonoscopy</p>
<p>14. Hoog CM, Bark LA, Brostrom O, Sjoqvist U. Capsule endoscopic findings correlate with fecal calprotectin and C-reactive protein in patients with suspected small-bowel Crohn's disease. <i>Scandinavian Journal of Gastroenterology</i>. 2014;49(9):1084-90.</p>	<p>Patients included have had negative colonoscopy and positive capsule endoscopy, small bowel only</p>
<p>15. Jensen MD, Nathan T, Rafaelsen SR, Kjeldsen J. Ileoscopy reduces the need for small bowel imaging in suspected Crohn's disease. <i>Danish Medical Journal</i>. 2012;59(9):A4491.</p>	<p>Comparison is ileocolonoscopy versus ileocolonoscopy plus small bowel capsule endoscopy</p>
<p>16. Joosen AMCP, Kok MB, Van Der Linden IJM, Bozkurt Z, Broos H, Van Pelt J, Van Heerde M, De Groot MJM. Analytical and clinical evaluation of faecal calprotectin. [Dutch] Analytische en klinische evaluatie van de bepaling van calprotectine in feces. <i>Nederlands Tijdschrift voor Klinische Chemie en Laboratoriumgeneeskunde</i>. 2013;38(4):196-201.</p>	<p>not sufficient data for 2x2 table and no demographics about patients, don't know whether adults and/or children, only reports PPV, NPV and concordance (TP+TN/N)</p>
<p>17. Kalantari H, Taheri A, Yaran M. Fecal calprotectin is a useful marker to diagnose ulcerative colitis from irritable bowel syndrome. <i>Advanced Biomedical Research</i>. 2015;4:85.</p>	<p>UC patients only, 21 patients with other colonoscopy findings (Crohn's colitis, microscopic colitis, collagenous colitis, or other) were excluded</p>
<p>18. Kok L, Elias SG, Witteman BJM, Goedhard JG, Muris JWM, Moons KGM, De Wit NJ. Diagnostic</p>	<p>Overlap with previous SR</p>

accuracy of point-of-care fecal calprotectin and immunochemical occult blood tests for diagnosis of organic bowel disease in primary care: The cost-effectiveness of a decision rule for abdominal complaints in primary care (CEDAR) study. <i>Clinical Chemistry</i> . 2012;58(6):989-98.	
19. Kopylov U, Yung DE, Engel T, Avni T, Battat R, Ben-Horin S, Plevris JN, Eliakim R, Koulaouzidis A. Fecal calprotectin for the prediction of small-bowel Crohn's disease by capsule endoscopy: a systematic review and meta-analysis. <i>European Journal of Gastroenterology & Hepatology</i> . 2016;28(10):1137-44.	Small bowel only, mix of patients with suspected and confirmed IBD, 5/7 studies investigated patients with negative endoscopy
20. Kotze LM, Nisihara RM, Marion SB, Cavassani MF, Kotze PG. FECAL CALPROTECTIN: levels for the ethiological diagnosis in Brazilian patients with gastrointestinal symptoms. <i>Arquivos de Gastroenterologia</i> . 2015;52(1):50-4.	No 2x2 data, number of IBD and IBS reported but FC levels only reported as means and medians, no measures of accuracy reported to derive 2x2 data from, data requested – no response
21. Koulaouzidis A, Sipponen T, Nemeth A, Makins R, Kopylov U, Nadler M, Giannakou A, Yung DE, Johansson GW, Bartzis L, Thorlacius H, Seidman EG, Eliakim R, Plevris JN, Toth E. Association Between Fecal Calprotectin Levels and Small-bowel Inflammation Score in Capsule Endoscopy: A Multicenter Retrospective Study. <i>Digestive Diseases & Sciences</i> . 2016;61(7):2033-40.	Small bowel only, mix of suspected and confirmed IBD
22. Lance S, White C. An audit on the appropriate use of faecal calprotectin testing within the Taranaki DHB: a case for a more discerning approach. <i>New Zealand Medical Journal</i> . 2015;128(1417):24-9.	Not sufficient information to derive 2x2 data
23. Li LQ, Zeng J, Wang S, Chen X, Jiang Z. Fecal calprotectin for diagnosis of inflammatory bowel disease: A meta-analysis. [Chinese]. <i>World Chinese Journal of Digestology</i> . 2016;24(31):4272-8.	Meta-analysis in Chinese, different question as included studies enrolled patients with confirmed diagnosis and healthy controls
24. Licata A, Randazzo C, Cappello M, Calvaruso V, Butera G, Florena AM, Peralta S, Camma C, Craxi A. Fecal calprotectin in clinical practice: a noninvasive screening tool for patients with chronic diarrhea. <i>Journal of Clinical</i>	Overlap with previous SR

Gastroenterology. 2012;46(6):504-8.	
25. Lozoya Angulo ME, de Las Heras Gomez I, Martinez Villanueva M, Noguera Velasco JA, Aviles Plaza F. Faecal calprotectin, an useful marker in discriminating between inflammatory bowel disease and functional gastrointestinal disorders. Gastroenterologia y Hepatologia. 2017;40(3):125-31.	Adults and children, results reported combined, author is unable to send results for adults separately
26. Manz M, Burri E, Rothen C, Tchanguizi N, Niederberger C, Rossi L, Beglinger C, Lehmann FS. Value of fecal calprotectin in the evaluation of patients with abdominal discomfort: An observational study. BMC Gastroenterology. 2012:5.	Overlap with previous SR
27. McFarlane M, Chambers S, Malik A, Lee B, Sung E, Nwokolo C, Waugh N, Arasaradnam R. Clinical outcomes at 12 months and risk of inflammatory bowel disease in patients with an intermediate raised fecal calprotectin: a 'real-world' view. BMJ Open. 2016;6(6):e011041.	Included patients with intermediate FC outcomes only
28. McFarlane M, Chambers S, Malik A, Lee B, Sung E, Nwokolo C, Waugh N, Arasaradnam R. Is NICE too optimistic about savings from normal faecal calprotectin results? Journal of Gastroenterology and Hepatology Research. 2016;5(1):1895-8.	Included patients with FC negative results only
29. Menees SB, Powell C, Kurlander J, Goel A, Chey WD. A meta-analysis of the utility of C-reactive protein, erythrocyte sedimentation rate, fecal calprotectin, and fecal lactoferrin to exclude inflammatory bowel disease in adults with IBS. American Journal of Gastroenterology. 2015;110(3):444-54.	Patients with confirmed IBD
30. Olsen PAS, Fossmark R, Qvigstad G. Fecal calprotectin in patients with suspected small bowel disease - a selection tool for small bowel capsule endoscopy? Scandinavian Journal of Gastroenterology. 2015;50(3):272-7.	Small bowel only, patients had negative / inconclusive bowel investigation
31. Osipenko MF, Livzan m A, Skalinskaia MI, Lialiukova EA. [Fecal calprotectin concentration in the differential diagnosis of bowel diseases]. Terapevticheskii Arkhiv. 2015;87(2):30-3.	Not sufficient information to derive 2x2 data

<p>32. Polewiczowska B, Ashraf W, Grainger S, Khan SR, Sandhu K. Negative faecal calprotectin (FC) is a strong predictor of negative small bowel capsule endoscopy. United European Gastroenterology Journal. 2013;1):A188-A9.</p>	<p>Small bowel only, not enough information on included patients</p>
<p>33. Rodriguez-Moranta F, Lobaton T, Rodriguez-Alonso L, Guardiola J. [Fecal calprotectin in the diagnosis of inflammatory bowel diseases]. Gastroenterologia y Hepatologia. 2013;36(6):400-6. Calprotectina fecal en el diagnostico de enfermedades inflamatorias.</p>	<p>Review</p>
<p>34. Samuel S, Rangunath K. Colonoscopy, inflammatory bowel disease. Endoscopy. 2013;45(4):289-91.</p>	<p>Review</p>
<p>35. Schulz C, Wex T, Arnim UV, Malfertheiner P. Validation of Two Calprotectin Rapid Tests in Daily Routine. Clinical Laboratory. 2016;62(7):1249-54.</p>	<p>Patients had confirmed IBD / IBS diagnosis, reference standard was ELISA for test comparison</p>
<p>36. Seenan JP, Thomson F, Rankin K, Smith K, Gaya DR. Are we exposing patients with a mildly elevated faecal calprotectin to unnecessary investigations? Frontline Gastroenterology. 2015;6(3):156-60.</p>	<p>Only patients with 100-200 microgramm/g FC characterised</p>
<p>37. Sipponen T, Haapamaki J, Savilahti E, Alfthan H, Hamalainen E, Rautiainen H, Koskenpato J, Nuutinen H, Farkkila M. Fecal calprotectin and S100A12 have low utility in prediction of small bowel Crohn's disease detected by wireless capsule endoscopy. Scandinavian Journal of Gastroenterology. 2012;47(7):778-84.</p>	<p>Overlap with previous SR - excluded as only investigation of small bowel</p>
<p>38. Smith LA, Gaya DR. Utility of faecal calprotectin analysis in adult inflammatory bowel disease. World Journal of Gastroenterology. 2012;18(46):6782-9.</p>	<p>Review</p>
<p>39. Sood R, Gracie DJ, Law GR, Ford AC. Systematic review with meta-analysis: the accuracy of diagnosing irritable bowel syndrome with symptoms, biomarkers and/or psychological markers. Alimentary Pharmacology & Therapeutics. 2015;42(5):491-503.</p>	<p>Not FC</p>
<p>40. Sydora MJ, Sydora BC, Fedorak RN. Validation of a point-of-care desk top device to quantitate fecal calprotectin and distinguish</p>	<p>Overlap with previous SR - excluded as patients had known diagnosis prior to the study</p>

inflammatory bowel disease from irritable bowel syndrome. <i>Journal of Crohns & Colitis</i> . 2012;6(2):207-14.	
41. Turvill J. High negative predictive value of a normal faecal calprotectin in patients with symptomatic intestinal disease. <i>Frontline Gastroenterology</i> . 2012;3(1):21-8.	Overlap with previous SR
42. von Arnim U, Wex T, Ganzert C, Schulz C, Malfertheiner P. Faecal calprotectin: a marker for clinical differentiation of microscopic colitis and irritable bowel syndrome. <i>Clinical & Experimental Gastroenterology</i> . 2016;9:97-103.	Spectrum bias - only included microscopic colitis patients (mixed active and in remission), no 2x2 data
43. Wagner M, Sjoberg K, Vigren L, Olesen M, Benoni C, Toth E, Carlson M. Elevated fecal levels of eosinophil granule proteins predict collagenous colitis in patients referred to colonoscopy due to chronic non-bloody diarrhea. <i>Scandinavian Journal of Gastroenterology</i> . 2016;51(7):835-41.	Not sufficient information to derive 2x2 data
44. Wang S, Shi H, Lu H, Wang F, Wei J, Wang Z, Liu C, Yuan B. Significance of fecal calprotectin in differential diagnosis of whole spectrum of digestive system diseases. [Chinese]. <i>Chinese Journal of Gastroenterology</i> . 2012;17(4):237-9.	Included patients had confirmed disease
45. Wang S, Wang Z, Shi H, Heng L, Juan W, Yuan B, Wu X, Wang F. Faecal calprotectin concentrations in gastrointestinal diseases. <i>Journal of International Medical Research</i> . 2013;41(4):1357-61.	no 2x2 data, sens / spec reported for 2 cut-offs from ROC, but no predictive value – request for further data failed, e-mail bounced back
46. Wassell J, Wallage M, Brewer E. Evaluation of the Quantum Blue rapid test for faecal calprotectin. <i>Annals of Clinical Biochemistry</i> . 2012;49(1):55-8.	Overlap with previous SR - excluded as no patient details reported

Table B3 Abstracts excluded from review

1. Alzoubaidi D, Asser L, Price T, Lithgo K, Housley D, Johnson MW. Is a false positive faecal calprotectin as false as you think? <i>Gut</i> . 2015;64:A238.	only considered FC+ patients with normal endoscopy
2. Astle VI, Lewis NR. Under-utilisation of faecal calprotectin to exclude ibd in patients with functional bowel disorders. <i>Gut</i> . 2014;63:A207-A8.	Cannot derive meaningful 2x2 data

<p>3. Atef S, Abdel Rahman S. The diagnostic performance of the combined use of fecal calprotectin and occult blood tests for screening of organic bowel disease in primary care setting. <i>Biochimica Clinica</i>. 2013;37:S658.</p>	<p>No test and cut off reported</p>
<p>4. Aujla UI, Hayee B, Sherwood R, Chung-Faye G. Predictive value of faecal calprotectin in patients undergoing colonoscopy: A cost-effective analysis. <i>United European Gastroenterology Journal</i>. 2013;1):A517.</p>	<p>Cannot derive meaningful 2x2 table</p>
<p>5. Caccaro R, Lollo G, Hatem G, Ugoni A, Buda A, D'Odorico A, Galeazzi F, D'Inca R, Savarino EV, Sturniolo GC. Are non-invasive markers of gastrointestinal disease predictors of enteropathy at small bowel capsule endoscopy? <i>United European Gastroenterology Journal</i>. 2014;1):A347-A8.</p>	<p>small bowel only in colonoscopy negative patients</p>
<p>6. Conroy S, Hale M, Cross S, Swallow K, Sidhu R, Sargur R, et al. Does the performance of faecal calprotectin testing in primary care differentiate patients with inflammatory bowel disease? <i>Gut</i>. 2015;64:A25.</p>	<p>Patients are children and adults no proportions reported</p>
<p>7. Demir OM, Ahmed Z, Logan RPH. Optimising the use of faecal calprotectin for early diagnosis of IBD in primary care. <i>Journal of Crohn's and Colitis</i>. 2013;7:S8-S9.</p>	<p>Cannot derive meaningful 2x2 data</p>
<p>8. Digby J, Steele RJC, Strachan JA, Mowat C. Stool tests can potentially rule out significant bowel disease in symptomatic patients in primary care. <i>United European Gastroenterology Journal</i>. 2014;1):A47.</p>	<p>Overlap of study population with Mowat 2016 suspected</p>
<p>9. Eccles J, Neely A, Lynch M, Ferguson CB, Morrison G. An overview of the impact of fecal calprotectin testing in the management of patients within the gastroenterology outpatient clinic in a general hospital. <i>Gastroenterology</i>. 2014;1):S-798.</p>	<p>Over 30% are children 15-18 years</p>
<p>10. Greig E, Gore S, Staveley K, Phillips I, Benneyworth R, Matull R, Thole S, Williams M. Delivering cost effective management for Irritable Bowel Syndrome (IBS) across Somerset. <i>Gut</i>. 2015;64:A192-A3.</p>	<p>Known cases of IBD and new cases included, proportions not reported</p>
<p>11. Hale MF, Drew K, McAlindon ME, Lobo AJ, Sidhu R. Faecal calprotectin in patients with suspected small bowel Crohn's disease: Correlation with small bowel capsule endoscopy. <i>Gut</i>. 2015;64:A83.</p>	<p>Colonoscopy negative patients only, investigation of small bowel only</p>

12. Hunt N, Allcock R, Sharma A, Myers M. Diagnostic performance of faecal calprotectin in primary care. <i>Gut</i> . 2014;63:A159.	Not enough information for 2x2 table
13. Kotze L, Nisihara R, Cavassani M, Valarini S, Kotze P. Fecal calprotectin levels for the ethiological diagnosis in Brazilian patients with gastrointestinal symptoms. <i>American Journal of Gastroenterology</i> . 2014;109:S500.	Superseded by full text, not enough data for 2x2 table, numbers of IBD and IBS don't add up to total, no numbers for FC levels or test accuracy measures
14. Kwok R, Peter F, Page BP, Ahmed T, Tay DXH, Woo ASJ, Constantinos A, Yip BCH, Sze KCP, Wee EWL. Diagnostic performance of faecal calprotectin in Singaporean patients. <i>Journal of Gastroenterology and Hepatology (Australia)</i> . 2016;31:212.	No 2x2 data, don't know definition of non-colitis group
15. Keny B, Gaikwad S, Parkar S, Dherai AJ, Shetty D, Desai D, Joshi A, Abraham P, Gupta T, Tester Ashavaid F. Fecal calprotectin-marker for inflammatory bowel disease. <i>Indian Journal of Clinical Biochemistry</i> . 2016;31 (1 Supplement 1):S59.	Cannot derive meaningful 2x2 table
16. Lee S, Borthwick H, Dhar A. Faecal calprotectin testing in primary and secondary care Are the current manufacturer's cut-off values clinically useful? <i>Journal of Crohn's and Colitis</i> . 2014;8:S155-S6.	Superseded by full text Lee 2013
17. Maheshwari P, Junagade P, Goulding C. Evaluation of the use of fecal calprotectin as a diagnostic aid for IBD in an Irish population. <i>Journal of Crohn's and Colitis</i> . 2014;8:S159.	Cannot derive meaningful 2x2 table
18. Malik A, Bowen D, Rees I. Usefulness of fecal calprotectin in clinical practice in a district general hospital. <i>Gut</i> . 2013;62:A65.	Duplicate abstract with Alrubaiy 2012
19. McFarlane M, Chambers S, Dhaliwal A, Patel A, Nwokolo C, Arasaradnam R. Six month clinical outcomes in patients with intermediate raised faecal calprotectin levels. <i>International Journal of Surgery</i> . 2015;23:S48.	intermediate FC levels only
20. McFarlane M, Dhaliwal A, Chambers S, Nwokolo C, Patel A, Arasaradnam R. Clinical outcomes in patients with intermediate raised faecal calprotectin levels. <i>United European Gastroenterology Journal</i> . 2014;1):A371-A2.	Not sufficient data for 2x2
21. Mishreki A, Bell H, Austin C, Sheppard S, Noblett S. Does a raised faecal calprotectin level correlate with a diagnosis of IBD on colonoscopic biopsies? <i>Colorectal Disease</i> . 2014;16:43.	No test and cut-off reported, clinical question not specified, differential diagnosis unknown

22. Mohammed N, Smale S. Positive calprotectin but negative investigations-what next? <i>Gut</i> . 2012;61:A236.	positive faecal calprotectin with normal endoscopic/radiological only
23. Moroni F, Winter JW, Morris AJ, Gaya DR. What is the clinical relevance of a mildly elevated faecal calprotectin detected in new referrals to the gastroenterology clinic? <i>Gut</i> . 2012;61:A78-A9.	only low and intermediate FC group included
24. Mukhtar A, Sivaramakrishnan N, Hassan F. Should the cut off values of faecal calprotectin for initiating further investigations be higher than current practice? <i>Gut</i> . 2016;65:A243-A4.	Only FC positive patients included
25. Oyaert M, Trouve C, Baert F, De Smet D, Langlois M, Vanpoucke H. Performance characteristics of faecal calprotectin testing for diagnosis of Inflammatory Bowel Disease. <i>Acta Clinica Belgica</i> . 2014;69:1-2.	Superseded by full publication
26. Pantaleoni S, Touscoz GA, Caviglia GP, Adriani A, Sguazzini C, Sapone N, Reggiani S, Rizzetto M, Astegiano M. Fecal calprotectin is an effective diagnostic tool that differentiates pathological from functional intestinal disorders. <i>Digestive and Liver Disease</i> . 2013;45:S104-S5.	patients with known diagnosis
27. Parker C, Lamb CA, Robinson M, Mansfield JC, Gunn M. Predicting inflammatory pathology at capsule enteroscopy: What is the utility of a raised faecal calprotectin? <i>Gut</i> . 2015;64:A74-A5.	excluded only FC positive patients
28. Patel KV, Zaman S, Fong S, Anderson SH. Comparison of video capsule endoscopy and faecal calprotectin as diagnostic tools in patients with abdominal symptoms suggestive of small bowel Crohn's Disease. <i>Journal of Crohn's and Colitis</i> . 2015;9:S153-S4.	small bowel only
29. Reed O, Doyle J, Murphy S. Faecal calprotectin to screen out IBS: Is it being used correctly? <i>Irish Journal of Medical Science</i> . 2015;184 (6 Supplement 1):S248.	incomplete data at time of analysis
30. Rouke JO, Dhaliwal A, Sagar V, Davies J, Milestone A. Use of faecal calprotectin in primary care to distinguish irritable bowel syndrome from inflammatory bowel disease. <i>Gut</i> . 2015;64:A86-A7. And 31. Dhaliwal A, J OR, Sagar V, Burdsall J, Ransford R, Milestone A. Use of faecal calprotectin pathway in primary care to distinguish irritable bowel syndrome from inflammatory bowel disease. <i>United European</i>	Identical, incomplete data available at time of analysis

Gastroenterology Journal. 2015;1):A237.	
32. Sartain S, Stone A. Faecal calprotectin levels of 50-200: Endoscopic findings and subsequent diagnoses. Gut. 2016;65:A197-A8.	Only intermediate FC levels
33. Shastri YM, Povse N, Stein J. Prospective evaluation of faecal tumour pyruvate kinase type M2 (M2-PK) in comparison to calprotectin in IBD patients. Journal of Crohn's and Colitis. 2013;7:S75-S6.	IBD monitoring
34. Shitrit AB, Braverman D, Paz K, Adar T, Koslowsky B, Goldin E. Faecal calprotectin and lactoferrin as biomarkers in patients undergoing capsule endoscopy. Gastroenterology. 2013;1):S424.	Small bowel only
35. Taylor N, Hills E, Sheen C, Al-Bahrani A, Grellier L. An audit of faecal calprotectin testing in suspected inflammatory bowel disease in the under 45's. Journal of Crohn's and Colitis. 2013;7:S131.	Clinical question not specified, non-IBD group unknown
36. Tomkins C, Zeino Z, Nwokolo C, Smith SC, Arasaradnam R. Faecal calprotectin analysis: Does the method matter? Gut. 2012;61:A173-A4.	No meaningful 2x2 data
37. Wu J, Bolton L, Chapman C, Chey CS, Harrison E, Kinderman H, Johnson H, Richards-Taylor A, Weaver S, McLaughlin S. Faecal calprotectin has an acceptable sensitivity for detecting small bowel crohn's disease: Results from real world clinical practice. Gut. 2016;65:A145.	Known diagnosis and small bowel only
38. Zaman S, Patel K, Goel R, Borrow DM, Anderson SH. Comparison of video capsule endoscopy and faecal calprotectin as diagnostic tools in patients with abdominal symptoms suggestive of small bowel crohn's disease. United European Gastroenterology Journal. 2013;1):A493.	Small bowel only
39. Zuhra N, Sartain S, Gordon J, Lloyd D. Capsule endoscopy in the investigation and management of small bowel crohn's disease. Gut. 2016;65:A167-A8.	Small bowel only, incomplete FC data

Table B4 Studies excluded from previous review (Waugh et al. 2013⁵⁸)

1. Ashorn S, Honkanen T, Kolho KL, Ashorn M, Valineva T, Wei B, et al. Fecal calprotectin levels and serological responses to microbial antigens among children and adolescents with inflammatory bowel disease. <i>Inflamm Bowel Dis</i> 2009;15:199–205.	Children
2. Basumani P, Bardhan K, Eyre R, Ellis R, The Rotherham Team. Faecal calprotectin: Rotherham experience (slide presentation). BSG Away day 28/06/2012 (accessed 19 July 2013). Unpublished	Published and included as Banerjee 2015
3. Canani RB, de Horatio LT, Terrin G, Romano MT, Miele E, Staiano A, et al. Combined use of noninvasive tests is useful in the initial diagnostic approach to a child with suspected inflammatory bowel disease. <i>JPGN</i> 2006;42:9–15.	Children
4. Diamanti A, Panetta F, Basso MS, Forgione A, Colistro F, Bracci F, et al. Diagnostic work-up of inflammatory bowel disease in children: the role of calprotectin assay. <i>Inflamm Bowel Dis</i> 2010;16:1926–30.	Children
5. Dolwani S, Metzner M, Wassell JJ, Yong A, Hawthorne AB. Diagnostic accuracy of faecal calprotectin estimation in prediction of abnormal small bowel radiology. <i>Aliment Pharmacol Ther.</i> 2004;20(6):615-21.	Small bowel only
6. Fagerberg UL, Loof L, Myrdal U, Hansson LO, Finkel Y. Colorectal inflammation is well predicted by fecal calprotectin in children with gastrointestinal symptoms. <i>JPGN</i> 2005;40:450–5.	Children
7. Henderson P, Casey A, Lawrence SJ, Kennedy NA, Kingstone K, Rogers P, et al. The diagnostic accuracy of fecal calprotectin during the investigation of suspected pediatric inflammatory bowel disease. <i>Am J Gastroenterol</i> 2012;107:941–9.	Children
8. Manz M, Burri E, Rothen C, Tchanguizi N, Niederberger C, Rossi L, et al. Value of fecal calprotectin in the evaluation of patients with abdominal discomfort: an observational study. <i>BMC Gastroenterol.</i> 2012;12:5.	Overlap with Burri 2013
9. Sidler MA, Leach ST, Day AS. Fecal S100A12 and fecal calprotectin as	Children

<p>noninvasive markers for inflammatory bowel disease in children. <i>Inflammatory Bowel Dis</i> 2008;14:359–66.</p>	
<p>10. Tomas AB, Vidal MV, Camps R. Fecal calprotectin as a biomarker to distinguish between organic and functional gastrointestinal disease. <i>Rev Esp Enferm Dig</i> 2007;99:689–93.</p>	<p>Children</p>
<p>11. Van de Vijver E, Schreuder AB, Crossen WR, Muller Kobold AC, van Rheenen PF. Safely ruling out inflammatory bowel disease in children and teenagers without referral for endoscopy [published online ahead of print 2012]. <i>Arch Dis Child</i> 2012;12:1014–18.</p>	<p>Children</p>

B5 TABLES OF STUDY CHARACTERISTICS

Table B5 Study characteristics - population

Study ID (author, year, country)	Aim	Population (N, setting, age, gender)	Study exclusions	Indication*	Clinical questions	Disease prevalence
Alrubaiy 2012 ¹⁴² ; UK Abstract	To evaluate the diagnostic value of FC as a non-invasive marker of bowel inflammation in routine out-patient gastroenterology clinic.	N=72 Setting: out-patients in gastroenterology clinic Mean age: Female: 44 years Male: 47 years Female: 44/72 (61%)	NR	Chronic diarrhoea, abdominal pain, abdominal distension, rectal bleeding	Inflammatory versus non-inflammatory disease	Inflammatory disease (IBD): 2/72 (2.7%)
Banerjee 2015 ⁵⁴ ; UK	To assess the role of FC in aiding differential diagnosis, maximising the numbers in whom IBD could be ruled out, making colonoscopy unnecessary, yet not miss anyone with the disease.	N=119 Setting: newly referred patients to gastrointestinal clinic Age: ≤40 years: 36% 41–60 years: 46% >60 years: 18% Female: 64/119 (54%)	Proven coeliac disease, pancreatic insufficiency	Diarrhoea of longer than 4 weeks duration	IBD versus IBS, Organic versus non-organic disease	IBD: 12/119 (10%)
Bharathi 2005 ²¹⁶ ; UK Abstract	To assess the negative predictive value of FC in excluding bowel pathology in young patients with suspected IBS.	N=58 Setting: NR Mean age: NR Female: NR	NR	Abdominal pain and/or loose stools	IBD versus IBS	IBD: 0/58
Boyd 2016 ⁶¹ ; UK Abstract	To assess an FC pilot pathway.	N=424 Setting: data from a primary care pilot FC pathway Mean age: NR Female: NR	Rejected samples	NR	IBD versus non-IBD	IBD: 23/424 (5.4%)
Burri 2013 ¹³⁵ ; Switzerland	To directly compare the test characteristics of FC and lactoferrin in their ability to detect organic intestinal disease	N=405 Setting: patients with abdominal discomfort referred for endoscopy	<18 years	abdominal discomfort	Organic versus other disease	Organic: 143/405 (35%)

Study ID (author, year, country)	Aim	Population (N, setting, age, gender)	Study exclusions	Indication*	Clinical questions	Disease prevalence
		Median age (range): 63 years (18–97) Female: 226 /405(56%)				
Carroccio 2003 ¹⁴³ , Italy	To evaluate the positive and negative predictive values of the FC assay in identifying the organic causes of chronic diarrhea.	N=70 Setting: patients referred for chronic diarrhea to an outpatient gastroenterology clinic Median age (range): 35 years (18–72) Female: 40/70 (57%)	Previous evaluation for chronic diarrhea, overt gastrointestinal bleeding, sigmoidoscopy or colonoscopy during the previous 2 years, familial adenomatous polyposis and hereditary nonpolyposis, colorectal cancer syndrome, pregnancy	Chronic diarrhea of unknown origin, lasting for more than 4 weeks, with or without abdominal pain	IBD versus IBS, IBD versus non-IBD, Organic versus non-organic disease	IBD: 11/70 (16%)
Caviglia 2014 ¹⁴⁴ , Italy	To assess FC concentration and evaluate the diagnostic accuracy in patients with and without bowel inflammation	N=66 Setting: first time referrals to outpatient gastroenterology clinic Mean age (range): 42 (18-78) years Female: 46/66 (70%)	Regular NSAIDs, known CRC or polyps, known acute gastrointestinal infection (diverticular disease)	Abdominal pain and/or altered bowel habit lasting at least 4 weeks	IBD versus IBS, Inflammatory versus non-inflammatory disease	IBD: 24/66 (36%)
Conroy 2017 ⁶³ , UK	To assess the diagnostic accuracy of FC in the detection of IBD in a primary care setting.	N=410 Setting: FC ordered in primary care Median age (range): 42 (16–91) years Female: 248/410 (60%)	< 16 years, known IBD, FC testing initiated in secondary care	NR	IBD versus non-IBD, Organic versus non-organic disease	IBD: 11/410 (2.7%)
Damms 2008 ¹³⁶ , Germany	To evaluate the diagnostic accuracy of the new calprotectin rapid test compared to an established ELISA test in detecting colonoscopy-proven intestinal inflammatory diseases	N=140 Setting: Patients referred to colonoscopy as in- or out-patients for clarification of lower GI symptoms Mean age (range): 58 (20–85) years Female: 78/144 (56%)	Known extraintestinal inflammatory diseases such as rheumatoid arthritis, chronic arthritis–sinusitis, urinary tract infection, patients on NSAID or anticoagulants	lower GI symptoms, altered bowel habit, weight loss, by reason of occult blood–rectal bleeding, search of tumor, resectional surgery, follow-up, or routine checkup	IBD versus functional disease, Organic versus non-organic disease	IBD: 18/140 (13%)

Study ID (author, year, country)	Aim	Population (N, setting, age, gender)	Study exclusions	Indication*	Clinical questions	Disease prevalence
	and colorectal malignancies.					
De Sloovere 2017 ¹³⁸ , Belgium	To evaluate two newly developed automated FC immunoassays in terms of analytical and diagnostic performance.	N=229 Setting: out- and inpatients with clinical suspicion of IBD and stool sample for FC measurement Median age (range): 34 (15–94) years Female: 150/229 (66%)	<15 years of age, absence of clinical data, unclear diagnosis or insufficient sample volume	Suspicion of IBD	IBD versus IBS, IBD versus non-IBD	IBD: 46/229 (20%)
Dhaliwal 2015 ¹⁰⁶ , UK	To determine the ability of FC to exclude IBD and optimal cut-offs for IBD in remission	N=311 Setting: FC samples in a secondary care setting Mean age: NR Female: 190/292 (65%)	NR	Suspected functional gastrointestinal disease (IBS) or IBD	IBD versus IBS	IBD: 148/311 (48%)
El-Badry 2010 ¹³⁴ , Egypt	To evaluate the sensitivity and diagnostic accuracy of FC assay at different cut-off values in discriminating between functional and organic gastrointestinal disorders	N=39 Setting: patients referred to hospital for diagnostic differentiation of functional or organic bowel disorders Mean age \pm SD: IBD: 39.4 \pm 15.9 years IBS: 34.3 \pm 15.75 years Organic (non-IBD): 40.2 \pm 14 years Female: 19 /39 (49%)	On NSAIDs, aspirin, and/or anticoagulants, concomitant non-gastrointestinal diseases, (rheumatoid arthritis, other connective tissue inflammatory diseases or liver cirrhosis)	abdominal pain, chronic diarrhoea, weight loss and/or anorectal bleeding	IBD versus IBS, IBD vs non-IBD	IBD: 9/39 (23%)
Garcia 2006 ²¹⁷ , Spain	To assess the usefulness of FC to predict the presence of pathological colonoscopy and to analyze its use to discriminate between	N=190 Setting: patients undergoing colonoscopy for medical indications Mean age \pm SD: Other: 59.5 \pm 6.0 years Polyps: 60.21 \pm 4.53 years	cardiopulmonary disease, kidney or liver disease, celiac disease, known malignancy	NR	Organic versus other disease	Organic disease: 73/190 (38%)

Study ID (author, year, country)	Aim	Population (N, setting, age, gender)	Study exclusions	Indication*	Clinical questions	Disease prevalence
	different organic diseases	CRC: 71.75 ± 6.64 years IBD: 35.84 ± 15.4 years Female: 98/190 (52%)				
Hogberg 2017 ⁶² ; Sweden	To assess the value of a point of care FIT and FC test in detecting CRC, high risk adenomas HRAs and IBD in primary care.	N=384 Setting: FC ordered in primary care Median age: 63.0 years Female: 241/384 (64.6%)	Patients declined, emergency admittance, FC not returned	Abdominal pain, change of bowel habits, diarrhoea, constipation, rectal bleeding, urgency, anaemia, incomplete stool evacuation, weight loss, Melaena, palpable mass	IBD versus non-IBD, Organic versus other disease	IBD: 10/373 (2.6%)
Jang 2016 ⁵² ; Korea	To evaluate the initial diagnostic ability of three faecal calprotectin kits in different bowel diseases	N=41 Setting: patients at hospital for workup of abdominal symptoms Mean age ± SD (range): CD patients: 22±5 (15–31) UC patients: 41±21 (19–74) Intestinal BD: 54±21 (39–69) IBS: 17±12 (5–31) Other colitis: 38±24 (11–69) Female: 13/41 (32%)	Incomplete colonoscopy, no faecal samples, CRC, history of bowel resection, uncertain diagnosis after colonoscopy, regular aspirin and/or NSAIDs	abdominal pain, changes in bowel habits, and/or anorectal bleeding	IBD versus IBS	IBD: 31/41 (76%)
Kennedy 2015 ⁵⁰ ; UK	To determine the most effective use of FC in the diagnosis of GI disease in patients with no prior known GI disease, at the first presentation to GI services.	N=895 Setting: Data of FC of patients presenting to gastrointestinal services Mean age: 33.1 years (IQR: 25.6–40.7) Female: 581/895 (64.9)	< 16 years or >50 years, Patients with FC tests from other hospitals, patients with confirmed GI diagnosis, on IBD treatment, insufficient detail available, severe intercurrent illness, on aspirin/NSAIDs, on corticosteroids and/or aminosaliculates	Bloody diarrhoea, watery diarrhoea, rectal bleeding, constipation, abdominal pain, weight loss, bloating, vomiting, dyspepsia, fatigue, possible extraintestinal manifestations	IBD versus functional disease, Organic versus non-organic disease	IBD: 91/895 (10%)

Study ID (author, year, country)	Aim	Population (N, setting, age, gender)	Study exclusions	Indication*	Clinical questions	Disease prevalence
Kok 2012 ⁴⁰ ; Netherlands	To quantify the diagnostic accuracy of 3 biomarker tests for the inclusion or exclusion of OBD in patients with persistent lower-abdomen complaints in primary care, who need colonoscopy referral	N=382 Setting: Referred patients to colonoscopy from primary care (recruitment was taking place at the GP practice or straight after registering for endoscopy Median age (range): 60 (18-91) years Female: 209/382 (54.7%)	<18 years old, unable to give informed consent, previously diagnosed with organic disease, or positive on triple feces test (TFT), not requiring endoscopy	Abdominal pain, rectal bleeding, diarrhoea, fever and weight loss	Organic versus other disease	Organic disease: 99/382 (26%)
Labaere 2014 ¹³¹ ; Belgium	To evaluate six FC tests and to determine their performance for diagnosis and follow-up of IBD patients	N=31 Setting: referred for colonoscopy to gastroenterology Mean age: 36 years (16-75) Female: 21/31 (68%)	Inconclusive histopathological and endoscopic evaluation, extremely watery consistency of the stool sample	Chronic diarrhoea and/or abdominal pain, iron deficiency anaemia, unexplained weight loss, or a family history of IBD	IBD versus non-IBD	IBD: 12/31 (39%)
Lee 2014 ²¹⁸ ; UK Abstract	To determine whether the manufacturer's cut-off for referral are clinically useful in making a positive diagnosis in patients presenting with chronic diarrhoea.	N=122 Setting: FC test results from primary and secondary care Mean age: NR Female: NR	NR	chronic diarrhoea	Organic versus non-organic disease	Organic: 23/122 (19%)
Li 2006 ³³ ; China	To assess the value of FC in differential diagnosis of IBS.	N=240 Setting: outpatients and hospitalised patients undergoing endoscopy Mean age \pm SD: IBS: 48 \pm 19years Inflammation: 42 \pm 16 years Female: 121/240 (50%)	Upper gastrointestinal symptoms, stomach/small intestinal disease, diseases of heart, lung, liver, kidney, nerve, mental disorder, alcoholic, pregnant women, drug addiction, long term use of NSAIDs, colorectal adenomas	NR	Inflammatory versus non-inflammatory disease	Inflammation: 60/240 (25%)

Study ID (author, year, country)	Aim	Population (N, setting, age, gender)	Study exclusions	Indication*	Clinical questions	Disease prevalence
Licata 2012 ²¹⁹ ; Italy	To assess the diagnostic performance of FC as a stool-screening biomarker for organic intestinal disease	N=346 Setting: outpatients with unexplained chronic, non-bloody diarrhoea referred for colonoscopy Median age (range): Inflammation: 41 (17-80) years No inflammation: 38 (18-87) years Female: 201/346 (58%)	Bleeding, known colorectal or gastric neoplasia, familial adenomatous polyposis and hereditary nonpolyposis colorectal cancer syndrome, history of colonic surgery, recent respiratory or urinary tract infection, acute infectious disease, pregnancy and alcohol abuse, on NSAIDs, aspirin, or anticoagulants within the previous month, CRC	Unexplained chronic (≥ 4 wk), non-bloody diarrhoea	Inflammatory versus non-inflammatory disease	Inflammation: 142/346 (41%)
Limburg 2000 ¹⁴⁷ ; USA	To assess and compare calprotectin and hemoglobin (Hb) as stool screening biomarkers for colorectal inflammation in unexplained chronic diarrhoea	N=128 Setting: Patients referred to colonoscopy Mean age (range): 57 years (21-85) Female: 84/128 (66%)	Abnormalities on GI x-rays, bleeding, GI endoscopy performed within preceding 2 weeks, CRC syndrome	Chronic diarrhea (≥ 4 wk duration) of unknown origin or chronic colitis of unknown activity	Inflammatory versus non-inflammatory disease	Inflammatory disease: 29/128 (23%)
Lizvan 2015 ¹³⁹ ; Russia	To assess the informational content of fecal noninvasive tests (calprotectin, transferrin, hemoglobin) in complex diagnosis of diseases of intestines	N=52 Setting: unclear Mean age (range): 38.6 years (18-49) Female: unclear	Anxiety symptoms, weight loss, consistent abdominal pain, fever, pathology in abdominal organs, anaemia, increased red blood cell sedimentation rate, abdominal surgery in the past 6 months prior to enrolment, NSAIDs, antibacterial agents within one month of enrolment	Abdominal pain or discomfort, changes in stool frequency over the last 3 months	Organic versus non-organic disease (IBS)	Organic disease: 36/52 (69%)
Mowat 2016 ¹³⁷ ; UK	To study the diagnostic accuracies of quantitative FHB and FC tests in	N=1031 Setting: referred patients from primary care	Faecal sample not suitable for analysis,	Rectal bleeding, anaemia, diarrhoea, altered bowel habit,	IBD versus non-IBD, Organic versus non-organic disease	IBD: 34/1031 (3.3%)

Study ID (author, year, country)	Aim	Population (N, setting, age, gender)	Study exclusions	Indication*	Clinical questions	Disease prevalence
	patients presenting to primary care with bowel symptoms	Median age (range): 64 years (16–90) Female: 564/1031 (55%)	returned samples outside the study period, known IBD, no colonoscopy, OGD only	abdominal pain and weight loss		
Otten 2008 ¹³³ , Netherlands	To evaluate the diagnostic accuracy of two new rapid calprotectin and lactoferrin fecal tests in assessing colonic inflammation in patients with chronic abdominal complaints.	N=139 Setting: patients referred for endoscopy or sigmoidoscopy by the GP (80%) or the gastroenterologist (20%) Mean age: IBS: 52.3 years IBD: 44.5 years Female: 61/114 (54%)	<18 years, patients with a history of colonic surgery and those with iron deficiency	Bloating, pain, change in defecation frequency or consistency, or blood and mucus in stool	IBD versus IBS	IBD 23/139 (17%)
Oyaert 2014 ¹⁴⁰ , Belgium	To compare two FC assays regarding their reliability in the use of diagnosis of IBD	N=183 Setting: out- and inpatients with suspicion of IBD and FC Median age (range): 32 years (14-89) Female: 104/183 (57%)	<14 years, previous IBD diagnosis, no colonoscopy, unclear diagnosis (indeterminate colitis), inability to collect enough faecal samples	Diarrhoea, mucous, bloody stools, weight loss, abdominal pain and cramping	IBD versus non-IBD	IBD: 51/183 (27.9%)
Oyaert 2017 ¹³⁰ , Belgium	To evaluate six different automated faecal calprotectin immunoassays for the diagnosis of IBD.	N=86 Setting: Patients in secondary care with suspicion of IBD and FC Median age (range): 35 years (14-94) Female: N=61/105 (58%)	<14 years, previous IBD diagnosis, no ileocolonoscopy, unclear diagnosis (e.g. indeterminate colitis), inability to collect enough faecal sample	Diarrhoea, mucous or bloody stools, weight loss, and abdominal pain and cramping	IBD versus non-IBD	IBD: 21/86 (24%)
Pavlidis 2013 ⁵¹ , UK	To assess the diagnostic performance of FC in routine general and the impact of different cut-offs on endoscopy and referral rates.	N=962 Setting: Suspected IBS patients in primary care Mean age \pm SD: 33 \pm 7 years Female: 577/962 (60%)	<18 and >45 years, previous diagnosis of IBD, missing clinical data, diarrhoea for <2 weeks	Persistent abdominal pain, bloating, alteration in stool form or frequency, passage of mucus	Organic versus non-organic disease	Organic disease: 94/962 (9.8%)
Rosenfeld 2016 ¹⁴⁸ ,	To evaluate the impact of FC test results for the	N=279	ischemic colitis, infectious enterocolitis or colorectal	Gastrointestinal symptoms	IBD versus IBS	IBD: 4/50 (8%)

Study ID (author, year, country)	Aim	Population (N, setting, age, gender)	Study exclusions	Indication*	Clinical questions	Disease prevalence
Canada	differentiation of GI symptoms as IBS or IBD.	Setting: patients for whom gastroenterologist considered FC to be an appropriate next step in diagnosis and/or management of symptoms Mean age \pm SD (range): 39.4 \pm 13.5 years (19-79) Female: NR	cancer, pregnancy, history of extensive bowel resection, ostomy, ileoanal pouch, NSAIDS, inability to collect a stool sample			
Schoepfer 2008 ¹³² , Switzerland	To determine the accuracy of 4 faecal markers in a patient group admitted for workup of GI symptoms.	N=136 Setting: Outpatients and inpatients admitted to colonoscopy \pm upper endoscopy Mean age (range): CD: 41 years (20–78) UC: 45 years (23–72) IBS: 40 years (20–79) Female: 54/94 (57%)	Incomplete ileocolonoscopy, microscopic colitis, infectious ileocolitis, colorectal cancer, colorectal polyps, unclear diagnosis (e.g., indeterminate colitis), inability to collect fecal samples, history of colorectal or small bowel surgery, regular intake of aspirin and/or an NSAID (≥ 2 tablets/week).	Abdominal pain, altered bowel habit, and/or anorectal bleeding	IBD versus IBS	IBD: 64/94 (68%)
Schroeder 2007 ⁴⁵ , Germany	To evaluate the clinical utility of FC to detect active gastrointestinal inflammation in patients suggestive of either having IBD or IBS.	N=76 Setting: patients referred for diagnostic clarification of chronic diarrhoea to tertiary referral practices Median age (range): CD: 40 (25–59) years UC: 38 (24–75) years IBS: 43 (20–72) years Female: 43/76 (57%)	Previous evaluation for chronic diarrhoea, rectal bleeding, sigmoidoscopy or colonoscopy during previous 2 months, colorectal cancer syndrome and pregnancy	Chronic diarrhoea	IBD versus IBS	IBD: 45/76 (59%)
Sharbatdaran 2018 ¹⁶⁸ Iran	To evaluate the level of fecal calprotectin in patients with IBD and	N=90	Chronic disease and illness that caused fever,	Clinical symptoms of colon diseases	IBD versus non-IBD	IBD: 45/90 (50%)

Study ID (author, year, country)	Aim	Population (N, setting, age, gender)	Study exclusions	Indication*	Clinical questions	Disease prevalence
	patients without inflammatory diseases of the colon	Setting: Referred to secondary care Mean age \pm SD: 34.69 \pm 10.42 years Female: 51/90 (56.7%)	Patients with calprotectin level of 50-200 μ g			
Shitrit 2007 ²²⁰ ; Israel	To assess the predictive value of faecal calprotectin in organic colonic disease.	N=72 Setting: patients referred to gastroenterology for colonoscopy Mean age \pm SD: 58 \pm 20 years Female: 35/72 (48%)	NSAIDs, concomitant serious illness, pregnancy, alcohol abuse, respiratory tract infection	NR	Organic versus non-organic disease	Organic disease: 28/72 (39%)
Sostres 2017 ¹⁴¹ ; Spain Abstract	To evaluate the diagnostic accuracy of FC in symptomatic patients referred for diagnostic colonoscopy.	N=171 Setting: Referred for diagnostic colonoscopy Median age (IQR): 62 (51-68) years Female: 73/171 (42.7)	NR	Rectal bleeding, change of bowel habits, anaemia	Organic versus non-organic disease	Organic disease: 37/171 (21.6%)
Tan 2016 ¹⁴⁶ ; UK Abstract	To evaluate the FC diagnostic referral pathway for primary care practitioners.	N=731 Setting: referred patients with FC tests performed in primary care Median age: 40 years Female: 461/731 (63%)	NR	NR	IBD versus non-IBD	IBD: 15/731 (2.1%)
Tibble 2002 ¹⁴⁵ ; UK	To determine if the use of FC is useful in differentiating between patients with organic and nonorganic disease.	N=602 Setting: patients referred to gastroenterology outpatient department by general practitioners Median age (range): 40 years (18–90) Female: 371/602 (62%)	Symptoms of oesophageal reflux, gastroesophageal pathology, or functional or isolated dyspepsia, known IBD, colorectal carcinoma, and serious cardiopulmonary, hepatic, renal, neurologic, and psychiatric disease	Abdominal pain, diarrhoea, constipation, anaemia, bleeding, weight loss	Organic disease versus other disease	Organic disease: 263/602 (44%)
Turvill 2012 ¹²⁴ ; UK	To determine the NPV of a normal faecal calprotectin in excluding organic intestinal disease in patients with intestinal symptoms referred	N=630 Setting: patients with intestinal symptoms referred from primary care to department of gastroenterology	Patients with fast track colorectal symptoms	Pain/discomfort, diarrhoea, constipation, bloating, bleeding	Organic versus non-organic disease	Organic disease: 109/630 (17%)

Study ID (author, year, country)	Aim	Population (N, setting, age, gender)	Study exclusions	Indication*	Clinical questions	Disease prevalence
	unselectively from primary care.	Mean age: Normal FC: 40.5 years Raised FC: 41 years Female: 442/630 (70%)				
Turvill 2016 ⁵⁵ ; UK	To evaluate the care pathway and recruit five primary care practices to conduct an FC implementation assessment.	N=262 Setting: patients presenting to primary care with new lower gastrointestinal symptoms Mean age \pm SD: 36.8 \pm 10.9 years Female: 183/262 (70%)	Patients in whom the GP suspected cancer	Diarrhoea, alternating bowel habit, pain, bloating, constipation	Organic versus non-organic disease	Organic disease: 26/262 (9.9%)

*Indication for either FC testing or for referral to colonoscopy if FC testing was not part of patient pathway

Table B6 Study characteristics – investigations

Study ID (author, year, country)	FC test and cut-off	Reference standard	Proportion without ref standard excluded	Proportion of analysis sample with reference standard	Definition target condition	Definition non-target condition
Alrubaiy 2014 ¹⁴² ; UK Abstract	NR 50 μ g/g	Colonoscopy +/- capsule endoscopy or follow-up at next clinic visit	None	28/54 (52%) colonoscopy +/- capsule endoscopy 26/54 (48%) follow-up at next clinic visit	IBD	No Bowel inflammation
Banerjee 2015 ⁵⁴ ; UK	IDK ELISA 8 μ g/g 25 μ g/g 50 μ g/g 75 μ g/g 100 μ g/g 150 μ g/g	Colonoscopy with histology	22/219 excluded from study	100%	IBD (excluding microscopic colitis)	Normal colonoscopy and histology
Bharathi 2005 ²¹⁶ ; UK Abstract	PhiCal ELISA 60 μ g/g	OGD, flexible sigmoidoscopy, colonoscopy, ultrasound, small bowel studies or clinical grounds	None	38/58 imaging 4/58 clinical grounds 16/58 unclear	IBD	IBS, non-intestinal, NSAIDS use, diverticulosis

Boyd 2016 ⁶¹ ; UK Abstract	NR 50µg/g 150µg/g	Investigation for referred not reported or median of 12 months (range 1-37) follow-up	None	Referred patients: NR Non-referred patients: follow-up	IBD (excluding microscopic colitis)	Gastroenteritis, microscopic colitis and IBS
Burri 2013 ¹³⁵ ; Switzerland	EK-CAL ELISA and PhiCal ELISA 50µg/g	Colonoscopy and histology +/- upper OGD	None recorded	70/405 with additional OGD	Organic: Esophagitis, erosive gastritis, gastric ulcer, gastric carcinoma, colitis (infectious colitis, CD, UC, diverticulitis, microscopic colitis), adenomatous polyp, CRC	Normal colonoscopy, hyperplastic polyp
Carroccio 2003 ¹⁴³ ; Italy	Calprest ELISA 50 µg/g 100 µg/g	Sigmoidoscopy or colonoscopy both with biopsy	None recorded	Sigmoidoscopy with biopsy if <40 years of age or colonoscopy with biopsy if >40	IBD (including microscopic colitis)	Colonoscopy / sigmoidoscopy negative and Manning criteria positive
Caviglia 2014 ¹⁴⁴ ; Italy	Calprest ELISA 50µg/g 100µg/g 150µg/g	Colonoscopy with biopsies	None reported	100%	IBD (including indeterminate colitis)	IBS
Conroy 2017 ⁶³ ; UK	IDK ELISA 50 µg/g 100 µg/g 125 µg/g 150 µg/g 250 µg/g	Colonoscopy, sigmoidoscopy, radiological features or small bowel video capsule endoscopy	None reported	146/410 (36%) colonoscopy (n=133) or sigmoidoscopy (n=13), 104/410 (25%) other investigation, 160 (39%) no investigation	IBD (including collagenous colitis, lymphocytic colitis)	Non-IBD (no further investigation and non-IBD organic conditions: CRC, colorectal adenomatous polyps, diverticulitis, appendicitis and diversion proctitis)
Damms 2008 ¹³⁶ ; Germany	EK-CAL ELISA 50 µg/g Prevista qualitative rapid	Colonoscopy with histology	None recorded	100%	IBD (UC and CD)	normal diagnostic findings
De Sloovere 2017 ¹³⁸ Belgium	EK-CAL ELISA 50µg/g 163µg/g 325µg/g	Ileocolonoscopy with histopathology or 'sum of all findings'	None	136/229 (59%), rest unclear	IBD (excluding microscopic colitis)	IBS

	fCal Turbo 50µg/g 81µg/g 86µg/g 280µg/g Diasorin CLIA 43µg/g 50µg/g 135µg/g					
Dhaliwal 2015 ¹⁰⁶ , UK	PhiCal ELISA 50 and 100µg/g	Endoscopy/histology and/or radiology	None recorded	NR	IBD excluding microscopic colitis	IBS fulfilling the Rome II criteria
El-Badry 2010 ¹³⁴ , Egypt	PhiCal ELISA 50 and 100 µg/g	Colonoscopy with biopsies	None recorded	100%	IBD	IBS following Rome II criteria
Garcia 2006 ²¹⁷ , Spain	Calprest 217mg/kg	Colonoscopy and histology	None recorded	100%	Organic disease: colon adenomas, CRC, IBD	Other: IBS, diverticulosis, NSAIDs or aspirin, PPI
Hogberg 2017 ⁶² , Sweden	Calpro ELISA 20µg/g 50µg/g 100µg/g	Bowel imaging (colonoscopy, CT colonography, double contrast barium enema) or 2-year follow-up	None	185/384 (48%) imaging, 199/384 (52%) 2-year follow-up	IBD including microscopic colitis and unspecific colitis	No or any other pathology (including CRC, HRA)
Jang 2016 ⁵² , Korea	Ridascreen ELISA Quantum-Blue rapid EliA FEIA 50µg/g	Colonoscopy	None recorded	100%	IBD including intestinal Behçet's disease	IBS
Kennedy 2015 ⁵⁰ , UK	PhiCal ELISA 20 µg/g 50 µg/g 70 µg/g 100 µg/g	Definitive organic diagnosis, full colonoscopy or minimum of 3 years follow-up	None	467/895 (52%) organic diagnosis or colonoscopy, 428/895 (48%) follow-up	IBD	Functional
Kok 2012 ⁴⁰ , Netherlands	EK-CAL ELISA Quantum-Blue rapid, 50 µg/g	Colonoscopy, sigmoidoscopy (89.9% with biopsy), other bowel examinations, 3 month follow-up	4/423 from study	Colonoscopy 351/423 (91.9%), sigmoidoscopy 21/423 (5.5%), other bowel	Organic disease: CRC, adenomas, IBD, and diverticulitis	Other: no structural abnormalities, diverticulosis, IBS, and haemorrhoids

				examinations 10/423 (2.6%) With histology in 89.9% Follow-up for inconclusive tests		
Labaere 2014 ¹³¹ ; Belgium	CalproLab ELISA 50µg/g Quantum-Blue rapid 50µg/g and 75µg/g Calfast rapid 70µg/g Calprest ELISA 50µg/g EliA FEIA 15µg/g and 50µg/g CerTest rapid 50µg/g	Ileocolonoscopies + biopsy	1/33excluded from study	100%	IBD	Non-organic conditions + non-IBD organic conditions (nonsteroidal anti-inflammatory drug enteropathy, adenoma, and infectious colitis)
Lee 2014 ²¹⁸ , UK Abstract	ELISA 60µg/g	Investigation in Gastroenterology clinic	None recorded	Patients with positive FC test	Organic disease: IBD, Diverticulosis, Colonic Polyps, Infective colitis and Chronic Pancreatitis	Functional
Li 2006 ³³ , China	PhiCal ELISA 50µg/g	Colonoscopy ± biopsy	None recorded	Patients with inflammation, unclear for patients with IBS	Inflammation: IBD, chronic infectious diarrhoea, intestinal tuberculosis, Behcet disease	IBS
Licata 2012 ²¹⁹ , Italy	Calprest ELISA 150 µg/g	Colonoscopy with biopsy	None recorded	100%	Inflammation: IBD, microscopic colitis, diverticulitis, polyps, ischemic colitis, nonspecific colitis, IBS or undiagnosed microscopic colitis	No inflammation: IBS, diverticulosis
Limburg 2000 ¹⁴⁷ , USA	PhiCal ELISA 100µg/g	Colonoscopy and histology	None recorded	100% (n=32 had no biopsy)	Inflammation: CD, UC, microscopic or collagenous colitis, peridiverticulitis, and eosinophilic colitis	normal mucosa, <1cm polyps
Lizvan 2015 ¹³⁹ , Russia	Quantum-Blue rapid 50µg/g	Colonoscopy	None recorded	100%	Organic disease	IBS

Mowat 2016 ¹³⁷ ; UK	EK-CAL ELISA 50µg/g 200µg/g	Lower endoscopy	163/1031 from analysis	100%	IBD	all non-IBD conditions and normal
Otten 2008 ¹³³ ; Netherlands	PhiCal ELISA 50µg/g CalDetect rapid 15µg/g 60µg/g	Colonoscopy or sigmoidoscopy biopsies were taken if necessary	2/144 from study	109 (96%) colonoscopy, and 5 (4%) sigmoidoscopy	IBD	IBS
Oyaert 2014 ¹⁴⁰ ; Belgium	Quantum-Blue rapid, ELiA ELISA 50µg/g	Ileocolonoscopy and histology	From study, proportion not reported	100%	IBD (excluding indeterminate colitis and microscopic colitis)	IBS, infectious colitis, oesophagitis, erosive gastritis, gastric ulcers, diverticulitis, microscopic colitis, CRC, hyperplastic polyps, adenomatous polyps, coeliac disease, arthritis
Oyaert 2017 ¹³⁰ ; Belgium	ELiA ELISA Diasorin CLIA Quanta Flash CLIA fCAL Turbo rapid Euroimmun ELISA Orgentec Alegria ELISA 50 µg/g and test specific optimal cut-off	Ileocolonoscopy and histology	From study, proportion not reported	100%	IBD (excluding microscopic colitis)	oesophagitis, erosive gastritis, gastric ulcers, diverticulitis, microscopic colitis, colorectal cancer, hyperplastic polyps, adenomatous polyps, spondylo-arthritis, (undifferentiated) arthritis, IBS
Pavlidis 2013 ⁵¹ ; UK	EK-CAL ELISA 50µg/g 100 µg/g 125 µg/g 150 µg/g 200 µg/g 249 µg/g	Colonoscopy, sigmoidoscopy, other investigations or 1 year follow-up	None	134/962 (14%) colonoscopy, 104/962 (11%) sigmoidoscopy, 130/962 (14%) other, 594/962 (62%) follow- up	Organic disease: IBD, Microscopic colitis, diversion colitis, infective diarrhoea, coeliac, NSAID, Alcohol related, Meckel's diverticulum, Rectal adenocarcinoma	Non-organic
Rosenfeld 2016 ¹⁴⁸ ; Canada	Quantum-Blue rapid, 100µg/g	Endoscopy, 26-months follow-up	None	8/41 FC negatives and 4/9 FC positives had endoscopy, remainder had 26 months follow-up	IBD including microscopic colitis	IBS

Schoepfer 2008 ¹³² ; Switzerland	PhiCal ELISA 50µg/g	Colonoscopy with biopsy	5/187 from study	100%	IBD	IBS according to Rome II criteria
Schroeder 2007 ⁴⁵ ; Germany	IDK ELISA 15µg/g	Colonoscopy with biopsy	None recorded	100%	IBD	IBS according to Rome II criteria
Sharbatdaran 2018 ¹⁶⁸ Iran	EK-CAL ELISA 127.65µg/g	Colonoscopy with biopsy	None	100%	IBD	Non-IBD inflammatory bowel diseases, IBS, no diagnosed disease
Shitrit 2007 ²²⁰ ; Israel	Calprest ELISA 150µg/g	Colonoscopy	None recorded	100%	Organic: IBD, carcinoma, polyps	Normal colonoscopy
Sostres 2017 ¹⁴¹ ; Spain Abstract	EliA FEIA 50µg/g	Colonic investigations (colonoscopy)	None	100%	CRC, advanced adenoma, IBD and angiodysplasia	NR
Tan 2016 ¹⁴⁶ ; UK Abstract	NR 50µg/g	Colonoscopies, gastroscopies, MRI of small bowel	None	NR for FC positive FC negatives: 58/95 colonoscopy 19/95 gastroscopy, 3/95 MRI of small bowel	IBD	NR
Tibble 2002 ¹⁴⁵ ; UK	ELISA 10mg/L=50µg/g	Barium enteroclysis/ enema and/or endoscopy/colonoscopy	None recorded	372/602 patients had a full colonoscopy	Crohn's disease, Celiac disease, Infective diarrhea, small bowel enteropathy (NSAID, Alcoholic, Radiation), Diabetic diarrhea, ulcerative colitis, Microscopic colitis, Collagenous colitis, Diverticular disease, cancer	IBS, IBS + nonulcer dyspepsia, IBS + other (lactose intolerance, angiodysplasia, hemorrhoids, and melanosis coli)
Turvill 2012 ¹²⁴ ; UK	PhiCal ELISA 50µg/g	Colonoscopy or barium enema ± histology, or various other investigations	None recorded	FC negative: 43% full evaluation of the colon by colonoscopy or barium enema and 60% had histology.	Organic disease: Crohn's disease, Ulcerative colitis, Indeterminate IBD, Microscopic colitis, NSAID enteropathy,	Non-organic: IBS

				FC positive: 75% underwent full colonic evaluation by colonoscopy or barium enema, supportive histology in 83%	Gastroenteritis, Diverticular disease, Bacterial overgrowth, Coeliac disease, Postoperative, Upper gastrointestinal bleed, Intussusception, Rectal polyp, Alcoholic enteropathy, Solitary rectal ulcer, Gastrinoma, Cholecystitis, Appendicitis, Bile salt malabsorption, Giardiasis, Chronic pancreatitis, Thyrotoxicosis, Small bowel bacterial overgrowth, Lactose intolerance, Sorbitol induced diarrhoea	
Turvill 2016 ⁵⁵ ; UK	EK-CAL ELISA 100µg/g 250µg/g	Colonoscopy, 6-week review of symptoms by GP or 6 months follow-up, hospital database	None	25% received colonoscopy, remainder received 6-week review of symptoms by GP or follow-up	Organic disease: IBD, non-specific inflammation, microscopic colitis, diverticular disease, gastroenteritis, coeliac disease, pancreatic insufficiency and a low-grade tubulovillous adenoma, haemorrhoids	IBS (incorporated all functional intestinal disease diagnoses) and non-enteric other diagnosis following a normal FC

B6 RISK OF BIAS AND APPLICABILITY CONCERN FOR EACH INCLUDED STUDY

Table B7 risk of bias and applicability concern for each included study

Study	RISK OF BIAS				APPLICABILITY CONCERNS		
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
Alrubaiy 2012	☹️	😊	☹️	☹️	☹️	😊	☹️
Banerjee 2015	😊	☹️	☹️	☹️	☹️	😊	😊
Bharathi 2005	☹️	☹️	☹️	☹️	☹️	😊	☹️
Boyd 2016	☹️	😊	☹️	☹️	😊	😊	☹️
Burri 2013	☹️	😊	☹️	☹️	☹️	😊	☹️
Carroccio 2003	😊	😊	☹️	☹️	☹️	😊	☹️
Caviglia 2014	☹️	😊	😊	☹️	☹️	😊	😊
Conroy 2017	😊	😊	☹️	☹️	☹️	😊	☹️
Damms 2008	☹️	😊	😊	☹️	☹️	😊	😊
De Sloovere 2017	☹️	😊	☹️	☹️	☹️	😊	☹️
Dhaliwal 2015	☹️	☹️	☹️	☹️	☹️	😊	☹️
El Badry 2010	☹️	☹️	☹️	😊	☹️	😊	😊
Garcia 2006	☹️	☹️	☹️	😊	☹️	😊	☹️
Hogberg 2017	😊	😊	☹️	☹️	☹️	😊	😊
Jang 2016	☹️	😊	☹️	☹️	☹️	😊	☹️
Kennedy 2015	😊	☹️	☹️	☹️	☹️	😊	☹️
Kok 2012	☹️	😊	😊	☹️	☹️	😊	☹️
Labaere 2014	☹️	☹️	😊	😊	☹️	😊	😊
Lee 2014	☹️	☹️	☹️	☹️	☹️	😊	☹️
Li 2006	☹️	☹️	☹️	☹️	☹️	😊	☹️
Licata 2012	😊	☹️	😊	☹️	☹️	😊	😊
Limburg 2000	☹️	😊	☹️	☹️	☹️	😊	☹️
Lizvan 2015	☹️	☹️	☹️	☹️	☹️	☹️	☹️
Mowat 2016	☹️	😊	☹️	☹️	☹️	😊	☹️
Otten 2008	☹️	😊	☹️	☹️	☹️	😊	☹️
Oyaert 2014	☹️	☹️	😊	☹️	☹️	😊	😊
Oyaert 2017	☹️	☹️	😊	☹️	☹️	😊	😊
Pavlidis 2013	☹️	😊	☹️	☹️	😊	😊	☹️
Rosenfeld 2016	☹️	😊	☹️	☹️	☹️	😊	☹️
Schoepfer 2008	☹️	😊	😊	😊	☹️	😊	😊
Schroeder 2007	😊	😊	😊	😊	☹️	😊	😊

Sharbatdaran 2018	☹️	☹️	😊	☹️	☹️	😊	😊
Shitrit 2007	😊	☹️	😊	😊	☹️	😊	☹️
Sostres 2017	😊	😊	😊	😊	☹️	😊	☹️
Tan 2016	😊	😊	☹️	☹️	😊	😊	☹️
Tibble 2002	😊	☹️	☹️	☹️	☹️	😊	☹️
Turvill 2012	😊	😊	☹️	☹️	☹️	😊	☹️
Turvill 2016	😊	😊	☹️	☹️	😊	😊	☹️

☹️ high 😊 unclear 😊 low

B7 PAIRED FOREST PLOTS OF SENSITIVITY AND SPECIFICITY OF 38 INCLUDED STUDIES BY CLINICAL QUESTION

Figure B1 presents the Forest plots of 14 studies reporting data for clinical questions IBD vs IBS and IBD vs functional disease. Figure B2 depicts the Forest plots of 16 studies reporting data for clinical questions IBD vs non-IBD and Inflammatory vs non-inflammatory intestinal disease. And Figure B3 shows the Forest plots of 19 studies reporting data for clinical questions organic intestinal disease vs non-organic intestinal disease and organic intestinal disease vs other intestinal conditions. Letters following the year indicate different 2x2 table data from the same study reference.

Study	TP	FP	FN	TN	test type	Clin question	cut-off in µg/g	Setting	test name	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Banerjee 2015 a	12	48	0	50	ELISA	IBD vs IBS	8.0	secondary	IDK	1.00 [0.74, 1.00]	0.51 [0.41, 0.61]		
Otten 2008b	23	5	0	86	POCT	IBD vs IBS	15.0	secondary	CalDetect	1.00 [0.85, 1.00]	0.95 [0.88, 0.98]		
Schroeder 2007	42	0	3	31	ELISA	IBD vs IBS	15.0	secondary	IDK	0.93 [0.82, 0.99]	1.00 [0.89, 1.00]		
Kennedy 2015a	90	289	1	277	ELISA	IBD vs functional	20.0	secondary	PhiCal	0.99 [0.94, 1.00]	0.49 [0.45, 0.53]		
Banerjee 2015 b	12	46	0	52	ELISA	IBD vs IBS	25.0	secondary	IDK	1.00 [0.74, 1.00]	0.53 [0.43, 0.63]		
De Sloovere 2017i	45	15	1	101	CLIA	IBD vs IBS	43.0	secondary	Diasorin	0.98 [0.88, 1.00]	0.87 [0.80, 0.93]		
Carroccio 2003a	11	8	0	32	ELISA	IBD vs IBS	50.0	secondary	Calprest	1.00 [0.72, 1.00]	0.80 [0.64, 0.91]		
Caviglia 2014 a	24	10	0	11	ELISA	IBD vs IBS	50.0	secondary	Calprest	1.00 [0.86, 1.00]	0.52 [0.30, 0.74]		
De Sloovere 2017h	43	14	3	102	CLIA	IBD vs IBS	50.0	secondary	Diasorin	0.93 [0.82, 0.99]	0.88 [0.81, 0.93]		
De Sloovere 2017a	46	55	0	61	ELISA	IBD vs IBS	50.0	secondary	EK-CAL	1.00 [0.92, 1.00]	0.53 [0.43, 0.62]		
Dammis 2008a	18	12	0	44	ELISA	IBD vs functional	50.0	secondary	EK-CAL	1.00 [0.81, 1.00]	0.79 [0.66, 0.88]		
Jang 2016a	30	1	1	4	FEIA	IBD vs IBS	50.0	secondary	EIA	0.97 [0.83, 1.00]	0.80 [0.28, 0.99]		
Banerjee 2015 c	12	39	0	59	ELISA	IBD vs IBS	50.0	secondary	IDK	1.00 [0.74, 1.00]	0.60 [0.50, 0.70]		
Dihalivai 2015	126	33	18	115	ELISA	IBD vs IBS	50.0	secondary	PhiCal	0.98 [0.81, 0.92]	0.79 [0.70, 0.84]		
El-Badry 2010a	9	3	0	17	ELISA	IBD vs IBS	50.0	secondary	PhiCal	1.00 [0.66, 1.00]	0.65 [0.62, 0.97]		
Otten 2008a	22	12	1	79	ELISA	IBD vs IBS	50.0	secondary	PhiCal	0.96 [0.78, 1.00]	0.87 [0.78, 0.93]		
Schoepfer 2008	53	0	11	30	ELISA	IBD vs IBS	50.0	secondary	PhiCal	0.83 [0.71, 0.91]	1.00 [0.88, 1.00]		
Kennedy 2015b	88	149	3	417	ELISA	IBD vs functional	50.0	secondary	PhiCal	0.97 [0.91, 0.99]	0.74 [0.70, 0.77]		
Dammis 2008b	16	11	2	45	POCT	IBD vs functional	50.0	secondary	Prevista	0.89 [0.65, 0.99]	0.80 [0.68, 0.90]		
Jang 2016b	31	1	0	4	POCT	IBD vs IBS	50.0	secondary	Quantum-Blue	1.00 [0.89, 1.00]	0.80 [0.28, 0.99]		
Jang 2016c	25	0	6	5	ELISA	IBD vs IBS	50.0	secondary	Ridascreen	0.81 [0.63, 0.93]	1.00 [0.48, 1.00]		
De Sloovere 2017d	46	30	0	86	PETIA	IBD vs IBS	50.0	secondary	fCal Turbo	1.00 [0.92, 1.00]	0.74 [0.65, 0.82]		
Otten 2008c	14	2	9	89	POCT	IBD vs IBS	60.0	secondary	CalDetect	0.61 [0.39, 0.80]	0.98 [0.92, 1.00]		
Bharathi 2005	0	12	0	46	ELISA	IBD vs IBS	60.0	secondary	PhiCal	Not estimable	0.79 [0.67, 0.89]		
Kennedy 2015c	88	112	3	454	ELISA	IBD vs functional	70.0	secondary	PhiCal	0.97 [0.91, 0.99]	0.80 [0.77, 0.83]		
Banerjee 2015 d	11	25	1	73	ELISA	IBD vs IBS	75.0	secondary	IDK	0.92 [0.62, 1.00]	0.74 [0.65, 0.83]		
De Sloovere 2017e	46	16	0	100	PETIA	IBD vs IBS	81.0	secondary	fCal Turbo	1.00 [0.92, 1.00]	0.86 [0.79, 0.92]		
De Sloovere 2017f	45	15	1	101	PETIA	IBD vs IBS	86.0	secondary	fCal Turbo	0.98 [0.88, 1.00]	0.87 [0.80, 0.93]		
Caviglia 2014 b	21	6	3	15	ELISA	IBD vs IBS	100.0	secondary	Calprest	0.88 [0.68, 0.97]	0.71 [0.48, 0.89]		
Banerjee 2015 e	11	18	1	80	ELISA	IBD vs IBS	100.0	secondary	IDK	0.92 [0.62, 1.00]	0.82 [0.73, 0.89]		
El-Badry 2010b	7	0	2	20	ELISA	IBD vs IBS	100.0	secondary	PhiCal	0.78 [0.40, 0.97]	1.00 [0.83, 1.00]		
Kennedy 2015d	87	74	4	482	ELISA	IBD vs functional	100.0	secondary	PhiCal	0.96 [0.89, 0.99]	0.87 [0.84, 0.90]		
Rosenfeld 2016	4	5	0	41	POCT	IBD vs IBS	100.0	secondary	Quantum-Blue	1.00 [0.40, 1.00]	0.89 [0.76, 0.96]		
De Sloovere 2017j	36	2	10	114	CLIA	IBD vs IBS	135.0	secondary	Diasorin	0.78 [0.64, 0.89]	0.98 [0.94, 1.00]		
Caviglia 2014 c	21	2	3	19	ELISA	IBD vs IBS	150.0	secondary	Calprest	0.98 [0.68, 0.97]	0.90 [0.70, 0.99]		
Banerjee 2015 f	10	14	2	84	ELISA	IBD vs IBS	150.0	secondary	IDK	0.83 [0.52, 0.98]	0.86 [0.77, 0.92]		
De Sloovere 2017b	45	11	1	105	ELISA	IBD vs IBS	163.0	secondary	EK-CAL	0.98 [0.88, 1.00]	0.91 [0.84, 0.95]		
De Sloovere 2017g	37	4	9	112	PETIA	IBD vs IBS	280.0	secondary	fCal Turbo	0.80 [0.66, 0.91]	0.97 [0.91, 0.99]		
De Sloovere 2017c	38	4	8	112	ELISA	IBD vs IBS	325.0	secondary	EK-CAL	0.83 [0.69, 0.92]	0.97 [0.91, 0.99]		

Figure B1 Forest plots of 14 studies reporting data for clinical questions IBD vs IBS and IBD vs functional disease

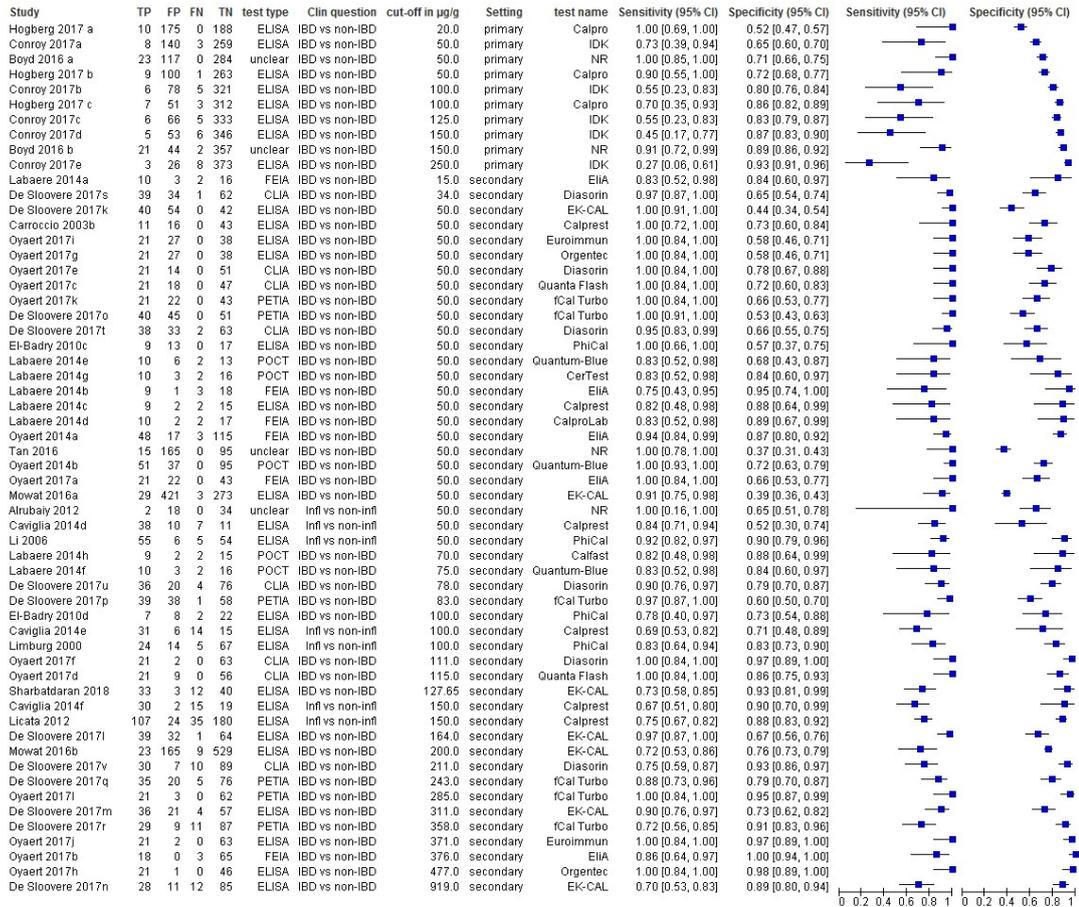
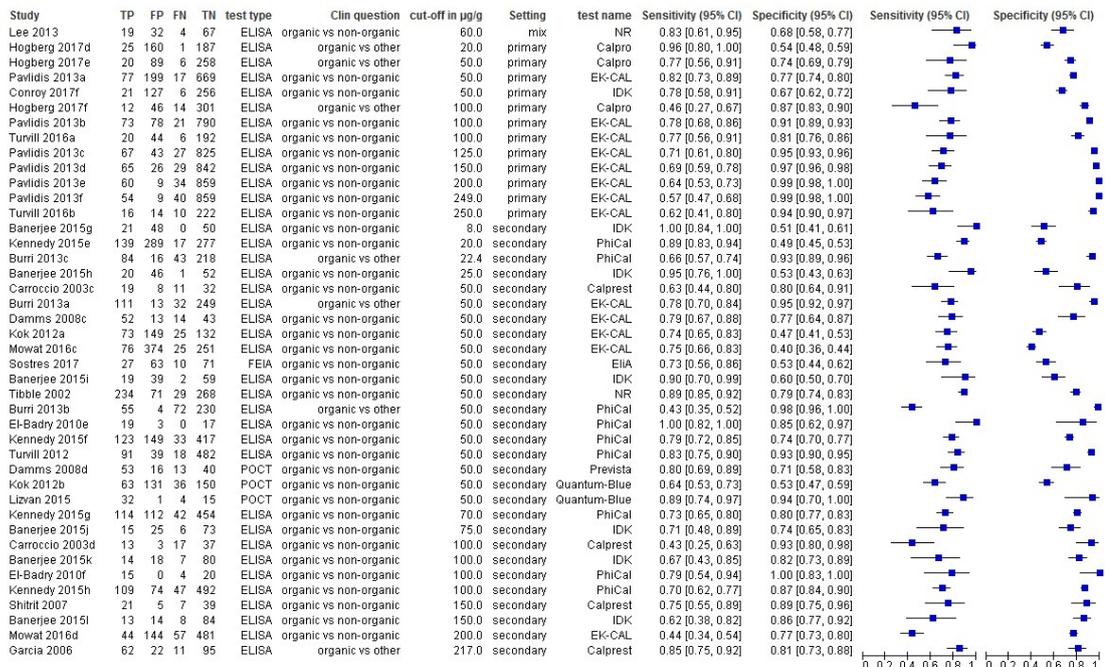


Figure B2 Forest plots of studies reporting data for clinical questions IBD vs non-IBD and Inflammatory vs non-inflammatory intestinal disease



B8 2X2 TABLE DATA OF STUDIES INCLUDED IN META-ANALYSIS BY TEST
(50µG/G THRESHOLD)

Table B8 2x2 table data of studies included in meta-analysis - EK-CAL (6 studies)

Study ID	Year of study	TP	FP	TN	FN	Test type	Test name
Burri 2013a	2013	111	13	249	32	ELISA	EK-CAL
Damms 2008c	2008	52	13	43	14	ELISA	EK-CAL
De Sloovere 2017a	2017	46	55	61	0	ELISA	EK-CAL
Kok 2012a	2012	73	149	132	25	ELISA	EK-CAL
Mowat 2016 a	2016	29	421	273	3	ELISA	EK-CAL
Pavlidis 2013a	2013	77	199	669	17	ELISA	EK-CAL

Table B9 2x2 table data of studies included in meta-analysis - ELiA (5 studies)

Study ID	Year of study	TP	FP	TN	FN	Test type	Test name
Jang 2016a	2016	30	1	4	1	FEIA	ELiA
Labaere 2014b	2014	9	1	18	3	FEIA	ELiA
Oyaert 2014a	2014	48	17	115	3	FEIA	ELiA
Oyaert 2017a	2017	21	22	43	0	FEIA	ELiA
Sostres 2017	2017	27	63	71	10	FEIA	ELiA

Table B10 2x2 table data of studies included in meta-analysis - PhiCal (8 studies)

Study ID	Year of study	TP	FP	TN	FN	Test type	Test name
Burri 2013b	2013	55	4	230	72	ELISA	PhiCal
Dhaliwal 2015	2015	126	33	115	18	ELISA	PhiCal
El-Badry 2010c	2010	9	13	17	0	ELISA	PhiCal
Kennedy 2015b	2015	88	149	417	3	ELISA	PhiCal
Li 2006	2006	55	6	54	5	ELISA	PhiCal
Otten 2008a	2008	22	12	79	1	ELISA	PhiCal
Schoepfer 2008	2008	53	0	30	11	ELISA	PhiCal
Turvill 2012	2012	91	39	482	18	ELISA	PhiCal

Table B11 2x2 table data of studies included in meta-analysis - Quantum-Blue (5 studies)

Study ID	Year of study	TP	FP	TN	FN	Test type	Test name
Jang 2016b	2016	31	1	4	0	POCT	Quantum-Blue
Kok 2012b	2012	63	131	150	36	POCT	Quantum-Blue
Labaere 2014e	2014	10	6	13	2	POCT	Quantum-Blue
Lizvan 2015	2015	32	1	15	4	POCT	Quantum-Blue
Oyaert 2014b	2014	51	37	95	0	POCT	Quantum-Blue

B9 2X2 TABLE DATA OF STUDIES INCLUDED IN META-ANALYSIS BY
CLINICAL QUESTION (50µG/G THRESHOLD)

Table B12 2X2 table data of studies included in meta-analysis - IBD versus IBS (11 studies)

Study ID	Year of study	TP	FP	TN	FN	Clinical question
Banerjee 2015	2015	12	39	59	0	IBD vs IBS
Carroccio 2003	2003	11	8	32	0	IBD vs IBS
Caviglia 2014	2014	24	10	11	0	IBD vs IBS
Damms 2008	2008	18	12	44	0	IBD vs functional
De Sloovere 2017	2017	46	55	61	0	IBD vs IBS
Dhaliwal 2015	2015	126	33	115	18	IBD vs IBS
El-Badry 2010	2010	9	3	17	0	IBD vs IBS
Jang 2016	2016	30	1	4	1	IBD vs IBS
Kennedy 2015	2015	88	149	417	3	IBD vs functional
Otten 2008	2008	22	12	79	1	IBD vs IBS
Schoepfer 2008	2008	53	0	30	11	IBD vs IBS

Table B13 2X2 table data of studies included in meta-analysis - IBD versus non-IBD (14 studies)

Study ID	Year of study	TP	FP	TN	FN	Clinical question
Alrubaiy 2012	2012	2	18	34	0	Infl vs non-infl
Boyd 2016	2016	23	117	284	0	IBD vs non-IBD
Carroccio 2003	2003	11	16	43	0	IBD vs non-IBD
Caviglia 2014	2014	38	10	11	7	Infl vs non-infl
Conroy 2017	2017	8	140	259	3	IBD vs non-IBD
De Sloovere 2017	2017	38	33	63	2	IBD vs non-IBD
El-Badry 2010	2010	9	13	17	0	IBD vs non-IBD
Hogberg 2017	2017	9	100	263	1	IBD vs non-IBD
Labaere 2014	2014	9	1	18	3	IBD vs non-IBD
Li 2006	2006	55	6	54	5	Infl vs non-infl
Mowat 2016	2016	29	421	273	3	IBD vs non-IBD
Oyaert 2014	2014	51	37	95	0	IBD vs non-IBD
Oyaert 2017	2017	21	18	47	0	IBD vs non-IBD
Tan 2016	2016	15	165	95	0	IBD vs non-IBD

Table B14 2X2 table data of studies included in meta-analysis - Organic versus non-organic (15 studies)

Study ID	Year of study	TP	FP	TN	FN	Clinical question
Banerjee 2015	2015	19	39	59	2	organic vs non-organic
Burri 2013	2013	111	13	249	32	organic vs other
Carroccio 2003	2003	19	8	32	11	organic vs non-organic
Conroy 2017	2017	21	127	256	6	organic vs non-organic
Damms 2008	2008	53	16	40	13	organic vs non-organic
El-Badry 2010	2010	19	3	17	0	organic vs non-organic
Hogberg 2017	2017	20	89	258	6	organic vs other
Kennedy 2015	2015	123	149	417	33	organic vs non-organic
Kok 2012	2012	63	131	150	36	organic vs non-organic
Lizvan 2015	2015	32	1	15	4	organic vs non-organic
Mowat 2016	2016	76	374	251	25	organic vs non-organic
Pavlidis 2013	2013	77	199	669	17	organic vs non-organic
Sostres 2017	2017	27	63	71	10	organic vs non-organic
Tibble 2002	2002	234	71	268	29	organic vs non-organic
Turvill 2012	2012	91	39	482	18	organic vs non-organic

B10 2X2 TABLE DATA AT 100µG/G THRESHOLD BY CLINICAL QUESTION

Table B15 2X2 table data of studies included in meta-analysis (100µg/g threshold) - IBD versus IBS (5 studies)

Study ID	Year of study	TP	FP	TN	FN	Clinical question	Threshold µg/g
Banerjee 2015	2015	11	18	80	1	IBD vs IBS	100
Caviglia 2014	2014	21	6	15	3	IBD vs IBS	100
El-Badry 2010	2010	7	0	20	2	IBD vs IBS	100
Kennedy 2015	2015	87	74	492	4	IBD vs functional	100
Rosenfeld 2016	2016	4	5	41	0	IBD vs IBS	100

Table B16 2X2 table data of studies included in meta-analysis (100µg/g threshold) - IBD versus non-IBD (5 studies)

Study ID	Year of study	TP	FP	TN	FN	Clinical question	Threshold µg/g
Caviglia 2014	2014	31	6	15	14	Infl vs non-infl	100
Conroy 2017	2017	6	78	321	5	IBD vs non-IBD	100
El-Badry 2010	2010	7	8	22	2	IBD vs non-IBD	100
Hogberg 2017	2017	7	51	312	3	IBD vs non-IBD	100
Limburg 2000	2000	24	14	67	5	Infl vs non-infl	100

Table B17 2X2 table data of studies included in meta-analysis (100µg/g threshold) - Organic versus non-organic (7 studies)

Study ID	Year of study	TP	FP	TN	FN	Clinical question	Threshold µg/g
Banerjee 2015	2015	14	18	80	7	organic vs non-organic	100
Carroccio 2003	2003	13	3	37	17	organic vs non-organic	100
El-Badry 2010	2010	15	0	20	4	organic vs non-organic	100
Hogberg 2017	2017	12	46	301	14	organic vs other	100
Kennedy 2015	2015	109	74	492	47	organic vs non-organic	100
Pavlidis 2013	2013	73	78	790	21	organic vs non-organic	100
Turvill 2016	2016	20	44	192	6	organic vs non-organic	100

B11 2X2 TABLE DATA AT 50µG/G THRESHOLD OF 28 STUDIES CONSIDERED FOR EXPLORATORY SENSITIVITY ANALYSIS

Table B18 28 Studies considered for exploratory sensitivity analysis

Study ID	Year of study	TP	FP	TN	FN
Alrubaiy 2012	2012	2	18	34	0
Banerjee 2015c*	2015	12	39	59	0
Banerjee 2015i	2015	19	39	59	2
Boyd 2016a	2016	23	117	284	0
Burri 2013a	2013	111	13	249	32
Burri 2013b	2013	55	4	230	72
Carroccio 2003a	2003	11	8	32	0
Carroccio 2003b	2003	11	16	43	0
Carroccio 2003c	2003	19	8	32	11
Caviglia 2014a	2014	24	10	11	0
Caviglia 2014d	2014	38	10	11	7
Conroy 2017a	2017	8	140	259	3
Conroy 2017f	2017	21	127	256	6
Damms 2008a	2008	18	12	44	0
Damms 2008b	2008	16	11	45	2
Damms 2008c	2008	52	13	43	14
Damms 2008d	2008	53	16	40	13
De Sloovere 2017a	2017	46	55	61	0
De Sloovere 2017d	2017	46	30	86	0
De Sloovere 2017h	2017	43	14	102	3
De Sloovere 2017k	2017	40	54	42	0

De Sloovere 2017o	2017	40	45	51	0
De Sloovere 2017t	2017	38	33	63	2
Dhaliwal 2015	2015	126	33	115	18
El-Badry 2010a	2010	9	3	17	0
El-Badry 2010c	2010	9	13	17	0
El-Badry 2010e	2010	19	3	17	0
Hogberg 2017b	2017	9	100	263	1
Hogberg 2017e	2017	20	89	258	6
Jang 2016a	2016	30	1	4	1
Jang 2016b	2016	31	1	4	0
Jang 2016c	2016	25	0	5	6
Kennedy 2015b	2015	88	149	417	3
Kennedy 2015f	2015	123	149	417	33
Kok 2012a	2012	73	149	132	25
Kok 2012b	2012	63	131	150	36
Labaere 2014b	2014	9	1	18	3
Labaere 2014c	2014	9	2	15	2
Labaere 2014d	2014	10	2	17	2
Labaere 2014e	2014	10	6	13	2
Labaere 2014g	2014	10	3	16	2
Li 2006	2006	55	6	54	5
Lizvan 2015	2015	32	1	15	4
Mowat 2016a	2016	29	421	273	3
Mowat 2016c	2016	76	374	251	25
Otten 2008a	2008	22	12	79	1
Oyaert 2014a	2014	48	17	115	3
Oyaert 2014b	2014	51	37	95	0
Oyaert 2017a	2017	21	22	43	0
Oyaert 2017c	2017	21	18	47	0
Oyaert 2017e	2017	21	14	51	0
Oyaert 2017g	2017	21	27	38	0
Oyaert 2017i	2017	21	27	38	0
Oyaert 2017k	2017	21	22	43	0
Pavlidis 2013a	2013	77	199	669	17
Schoepfer 2008	2008	53	0	30	11
Sostres 2017	2017	27	63	71	10
Tan 2016	2016	15	165	95	0
Tibble 2002	2002	234	71	268	29
Turvill 2012	2012	91	39	482	18

*Letters indicate different 2x2 table data from the same study reference

B12 SETTING CRITERIA FOR INCLUSION OF PATIENTS INTO THE TEST ACCURACY STUDY

Exploring units of measurements of FC test results

I explored whether the unit mmol/l was used incorrectly and whether it could be replaced by $\mu\text{g/g}$ by comparing the distribution of test results in mmol/l with the distribution of test results reported in $\mu\text{g/g}$.

Strictly, FC test results recorded in mg/kg units of measurement are mathematically equivalent to those recorded in $\mu\text{g/g}$. However, to ensure there were no inconsistencies I explored whether these units can be used interchangeably by testing whether test results reported in different units were derived from the same distribution.

I compared distributions using the two-sample Kolmogorov-Smirnov test. This nonparametric hypothesis test can be used to test the agreement between two sample distributions. It does that by, first, computing the observed cumulative distribution functions of the two samples and, second, evaluating their difference. The Kolmogorov-Smirnov statistic (D_{stat}) is the maximum absolute difference of the two distribution functions. The test statistic is compared with a critical value (D_{critical}). At $D_{\text{stat}} > D_{\text{critical}}$ the null hypothesis that results reported with the two units can be treated as equivalent was rejected (at significance level $\alpha=0.05$).

A total number of 7,085 FC tests had a numeric result recorded in $\mu\text{g/g}$, 562 tests had a numeric test result reported in mg/kg and 261 in mmol/l. The histograms in Figure B4 show the distribution of test results by unit of measurement from 0 to 500. The histograms suggest that results reported in different units may have been drawn from the same distribution. Table B19 portrays the corresponding summary statistics for the numeric results by unit of measurement. I would have expected the median to be similar for test results with different units of measurement if they were drawn from the same distribution. However, the median for results reported in $\mu\text{g/g}$ is twice as large as the median for results reported in mg/kg (33 $\mu\text{g/g}$ versus 15mg/kg). Furthermore, the two-sample Kolmogorov-Smirnov test to test whether the above histograms with units $\mu\text{g/g}$ and mg/kg are sampled from the same distribution ($D_{\text{stat}} = 0.4965$, p-value <0.0001 , $D_{\text{critical}} = 0.05960001$) is statistically significant suggesting that the samples may not have been taken from the same distribution.

Similarly, the data suggest that the results reported in $\mu\text{g/g}$ (median $33\mu\text{g/g}$) and mmol/l (median 84mmol/l , $D_{\text{stat}} = 0.23915$, $p\text{-value} < 0.0001$, $D_{\text{critical}} = 0.08571843$) may not stem from the same distribution. This means that the units mg/kg and mmol/l could be genuinely different from $\mu\text{g/g}$. Alternatively, the units could be the same but their distributions differ because a single laboratory which uses the units may serve a certain significantly different subset of the population.

As a result I am not confident to replace mg/kg with $\mu\text{g/g}$ or to assume that mmol/l was used erroneously. Without conversion, the results reported in mmol/l cannot be interpreted the same as results reported in $\mu\text{g/g}$. To convert mmol/l into $\mu\text{g/g}$ I would need the molecular mass of calprotectin and the amount of stool sample used in the analysis for FC level using the formula: $\text{number of moles} = \text{mass of sample} / \text{molecular weight}$. However, I do not know the standardised amount of stool sample used in the analysers and whether laboratories use the same analysers and could not convert mmol/l into $\mu\text{g/g}$.

Table B19 Common test statistics for distributions of numeric FC test results reported as $\mu\text{g/g}$, mg/kg and mmol/l

Unit	N	Mean	Median	IQR	Range
$\mu\text{g/g}$	7085	124.5481	33	78	4.9-12064
mg/kg	562	79.22313	15	20.75	3.4-2756
mmol/l	261	258.0253	84	234	6-2500

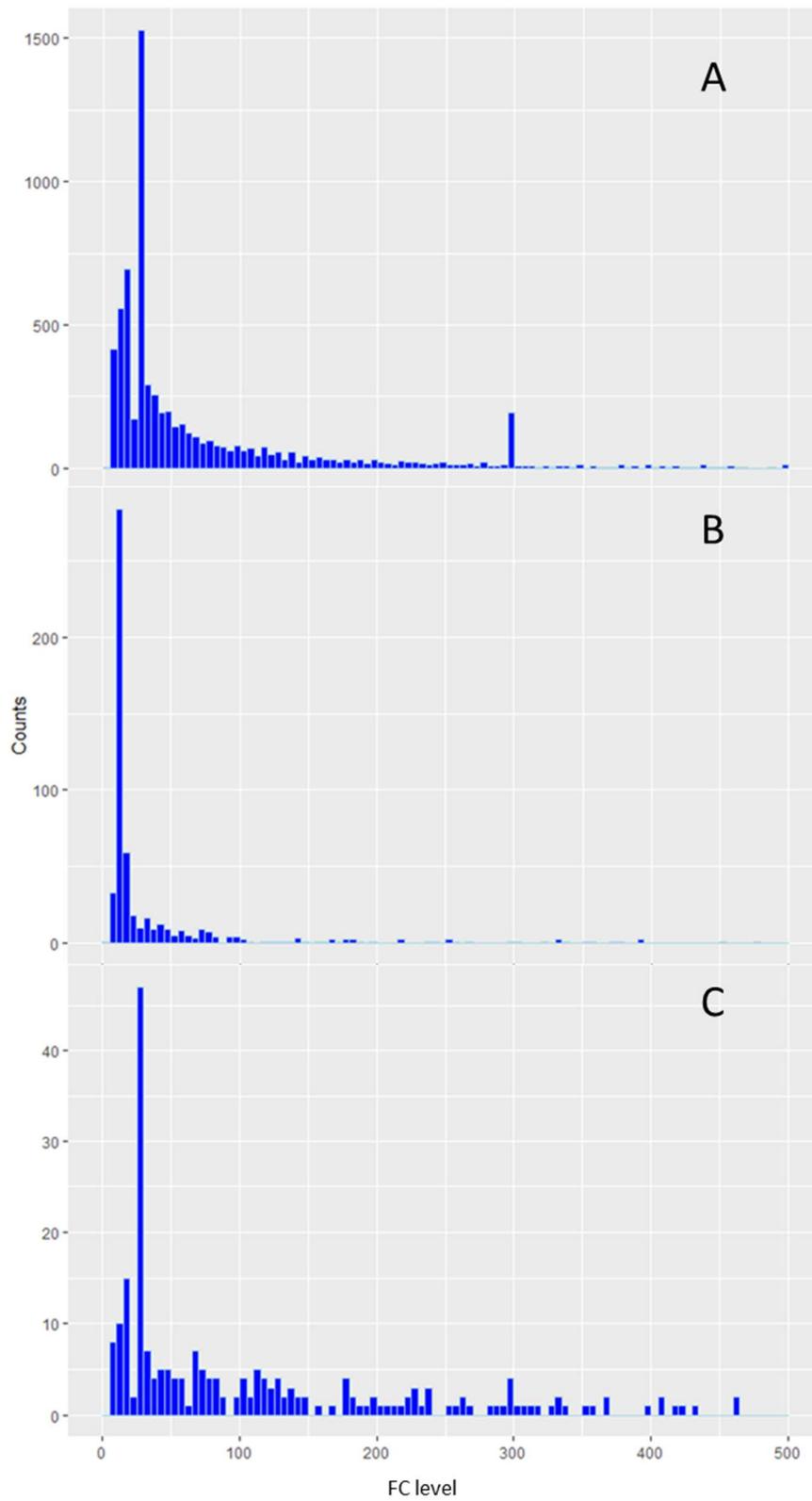


Figure B4 Distribution of numeric FC test results recorded in A) $\mu\text{g/g}$, B) mg/kg and C) mmol/l

Exploring operators of numeric FC test results

The histogram of FC levels in $\mu\text{g/g}$ in Figure B4 (A) reveals an unexpected peak at $300\mu\text{g/g}$ which may represent a common threshold that is used for reporting. To explore the peak I investigated the numeric results of $300\mu\text{g/g}$ by operator. 194/7,085 numeric results had an FC value of $300\mu\text{g/g}$. These 194 results were recorded by a subset of 37/478 practices. Of the 194 results 187 had a preceding operator “>” while 7 were preceded by “=”. This suggests that at least one laboratory considered FC values above $300\mu\text{g/g}$ to be no different in terms of their interpretation. This prompted me to investigate all 7,085 numeric results by operator to determine the proportion of results by preceding operator and how I should interpret those in ROC analyses. This investigation was restricted to FC tests with a numeric test result recorded in $\mu\text{g/g}$ and an appropriate operator (N=7,085).

Numeric values with operator “>”

Out of 7,085 numeric results, 273 were reported using the operator “>”. Except for one value of $>30\mu\text{g/g}$, results reported with the operator “>”, used the threshold of $300\mu\text{g/g}$ and above with the majority using the $300\mu\text{g/g}$ threshold (Figure B5, Table B20). As this threshold is greater than the 50 and $100\mu\text{g/g}$ thresholds used in my analyses all values will be greater than $100\mu\text{g/g}$ and can be included as positive at both thresholds. I decided to exclude the single value reported as $>30\mu\text{g/g}$ as it is unknown what the true numeric value is and whether this would be positive at the 50 and $100\mu\text{g/g}$ thresholds.

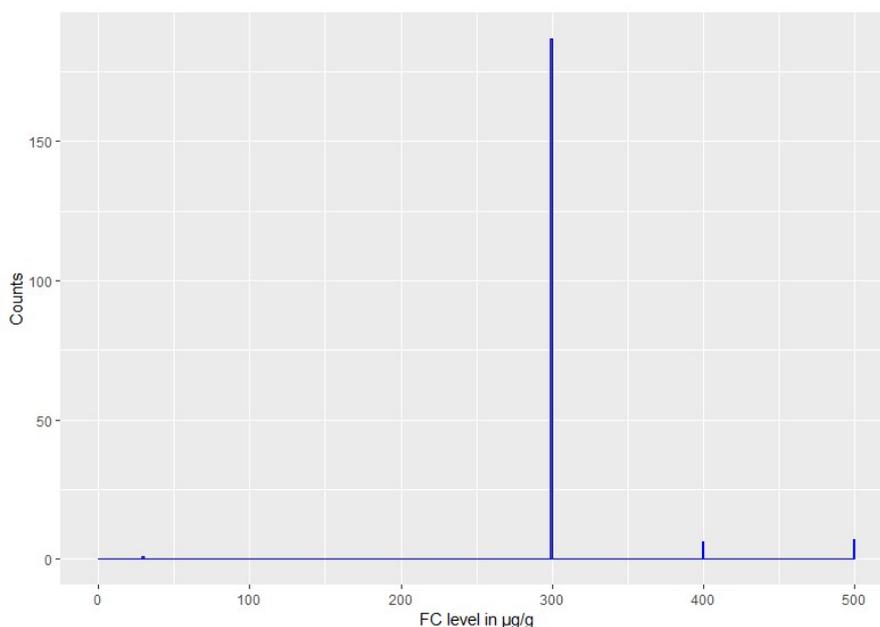


Figure B5 Distribution of numeric FC results in $\mu\text{g/g}$ preceded by the “>” operator

Table B20 Summary statistics of the distribution of numeric FC results in $\mu\text{g/g}$ preceded by the “>” operator

Summary statistic	Minimum	1 st Quartile	Median	Mean	3 rd Quartile	Maximum
FC level	30.0	300.0	300.0	585.5	600.0	9000.0

Numeric values with operator “<”

2262/7085 numeric results were reported with the operator “<” and one result with “ \leq ” which is included in this group. All results reported using the operator “<” use the threshold $33\mu\text{g/g}$ or smaller (Figure B6, Table B21). This means all actual results are smaller than 50 and $100\mu\text{g/g}$ and can be included in the analysis of test accuracy at both thresholds.

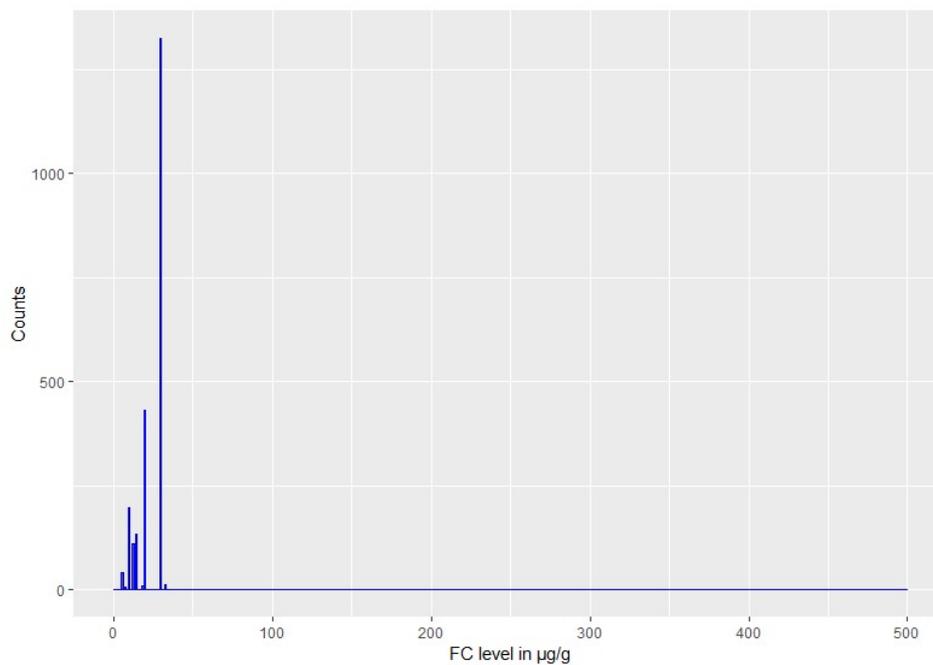


Figure B6 Distribution of numeric FC results in $\mu\text{g/g}$ preceded by the “<” operator

Table B21 Summary statistics of the distribution of numeric FC results in $\mu\text{g/g}$ preceded by the “<” operator

Summary statistic	Minimum	1 st Quartile	Median	Mean	3 rd Quartile	Maximum
FC level	5.70	19.50	30.00	24.11	30.00	33.00

Numeric values with operator “=”

The histogram in Figure B7 of the 4,548 results reported with the operator “=” shows that as expected the peak at 300µg/g has disappeared. Interestingly, the median has also increased considerably (59 versus 33µg/g) by disregarding the great proportion of results reported as <33µg/g and below (Table B22).

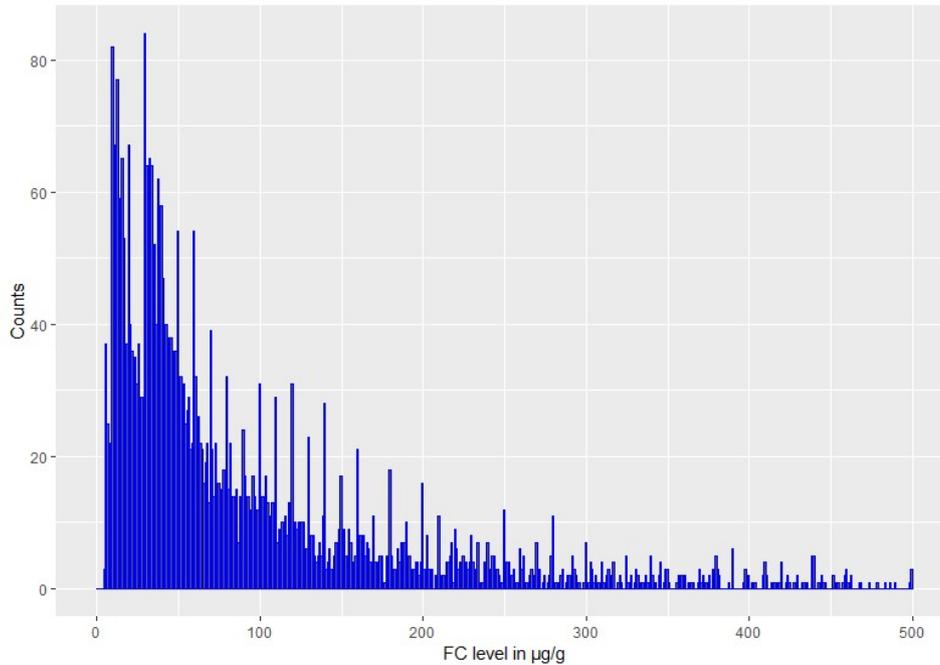


Figure B7 Distribution of numeric FC results in µg/g preceded by the “=” operator

Table B22 Summary statistics of the distribution of numeric FC results in µg/g preceded by the “=” operator

Summary statistic	Minimum	1 st Quartile	Median	Mean	3 rd Quartile	Maximum
FC level	4.9	31.0	59.0	146.9	133.0	12064.0

B13 DESCRIPTION OF THIN VARIABLES RELATING INFORMATION ON THE FC TEST RESULT

Table B23 THIN variables relating information on the FC test result

THIN variable relating to FC test result	Definition and information
Medcode description	Medcode descriptions describe the medical Read codes that GPs have available to record an FC test in a patient's clinical record. There are 6 different codes available two of which convey a qualitative outcome of the FC test undertaken. However, the majority of medcodes used to record an FC test ("Calprotectin level" and "Faecal calprotectin content") did not convey information on test outcome (17,317/17,466 [99.1%]).
Operator	An operator often preceded the numeric result. In the majority of cases the operator was an equal sign (7,864/11,324 [69.4%]). However, systems also made use of the operator less than '<' and to a lesser degree greater than '>'. Numeric test results with operators other than '=' cannot be taken at face value. This needs to be considered in analyses involving test accuracy measures at a threshold greater or smaller than the value recorded as the exact numeric value for these tests is unknown.
Numeric result	The numeric result represents the measured level of faecal calprotectin in the stool sample. The measurement for FC ranged from 0 to 15,000 (median 32.0).
Units	Units of measurement were reported 9,899 times. Generally, the unit was part of a numeric value (8,152/9,899 [82.4%]). The most commonly unit used in the literature and by laboratories is µg/g. This was also the most frequently recorded unit in the THIN database. However, there were an additional 14 units that were used to record FC test results in the database which suggests poor reporting. The second and third most frequently units used were mmol/l and mg/kg. The former is uncharacteristic for protein levels in stool samples but commonly used in more routine blood test including serum cholesterol and blood sugar. One explanation of the use of this unit is the incorrect set up of the laboratories' reporting system.
Significance	The significance variable provides qualitative outcomes in form of information provided by laboratories to GPs which supports the action to be taken in response to the test result. This information is only useful for a test accuracy study if the threshold used for classification is known.
Upper reference range	The upper reference range provides information on the threshold for the interpretation of numeric and qualitative test results and was reported for 12,763 FC tests. The majority of FC test results were interpreted at the threshold of 50µg/g. I am uncertain about the meaning of an upper reference range of 0. This may indicate that this information should also be viewed with caution. Information on the upper reference range can help in the interpretation of a qualitative FC test outcome recorded as a medcode or significance level.

B14 ASSOCIATION OF VARIABLES WITH TYPE OF DIAGNOSIS BY FC TEST OUTCOME

Table B24 Association of symptoms, referral, further testing and comorbidities with type of diagnosis by FC test outcome based on 2,414 positives and 3,556 negative FC tests

Characteristic	FC positives				FC negatives				OR (95% CI) Cases = IBD, non-cases = no IBD
	IBD within 6 months (n=187)	IBS within 6 months (n=100)	No diagnosis within 6 months (n=2,096)	No diagnosis after FC (n=1,470)	IBD within 6 months (n=14)	IBS within 6 months (n=262)	No diagnosis within 6 months (n=3,277)	No diagnosis after FC (n=2,284)	
Referral within 6 weeks of FC test									
Referral n (%)	131 (70.05)	66 (66.00)	1,248 (59.54)	870 (59.18)	10 (71.43)	130 (49.62)	1,467 (44.77)	1020 (44.66)	2.13 (1.64 to 2.75)
Comorbidities recorded at any time									
Arthritis n (%)	3 (1.60)	0 (0.00)	23 (1.10)	18 (1.22)	0 (0.00)	1 (0.38)	16 (0.49)	14 (0.61)	1.89 (0.65 to 5.47)
Asthma n (%)	40 (21.39)	22 (22.00)	327 (15.60)	237 (16.12)	3 (21.43)	55 (20.99)	541 (16.51)	381 (16.68)	1 (0.72 to 1.4)
Depression n (%)	35 (18.72)	29 (29.00)	586 (27.96)	400 (27.21)	5 (35.71)	72 (27.48)	882 (26.91)	647 (28.32)	0.71 (0.52 to 0.98)
Dyspepsia n (%)	22 (11.76)	16 (16.00)	358 (17.08)	277 (18.84)	1 (7.14)	26 (9.92)	373 (11.38)	270 (11.82)	0.97 (0.67 to 1.41)
Abdominal symptoms recorded within 1 year prior to FC test									
Abdominal pain n (%)	37 (19.79)	36 (36.00)	533 (25.43)	331 (22.52)	4 (28.57)	77 (29.39)	898 (27.40)	594 (26.01)	0.69 (0.5 to 0.95)
Bloating n (%)	4 (2.14)	5 (5.00)	59 (2.81)	43 (2.93)	1 (7.14)	8 (3.05)	152 (4.64)	100 (4.38)	0.51 (0.24 to 1.07)
Bowel habit n (%)	11 (5.88)	3 (3.00)	131 (6.25)	94 (6.39)	0 (0.00)	14 (5.34)	228 (6.96)	180 (7.88)	0.73 (0.41 to 1.29)
Constipation n (%)	5 (2.67)	2 (2.00)	45 (2.15)	37 (2.52)	0 (0.00)	11 (4.20)	67 (2.04)	50 (2.19)	0.83 (0.43 to 1.6)

Diarrhoea n (%)	79 (42.25)	21 (21.00)	589 (28.10)	402 (27.35)	6 (42.86)	58 (22.14)	683 (20.84)	480 (21.02)	1.63 (1.21 to 2.2)
NICE symptoms n (%)	18 (9.63)	10 (10.00)	229 (10.93)	127 (8.64)	2 (14.29)	25 (9.54)	286 (8.73)	175 (7.66)	1.33 (0.99 to 1.78)
Subsequent FC tests within 6 months of initial FC test									
1 further test n (%), [of which prior to diagnosis]	16 (8.56), [8]	18 (18.00), [14]	311 (14.84), [NA]	214 (14.56), [NA]	2 (14.29), [2]	13 (5.00), [5]	92 (2.81), [NA]	64 (2.80), [NA]	1.38 (0.93 to 2.07)
2 further tests n (%)	5 (2.67), [2]	1 (1.00), [1]	34 (1.62), [NA]	22 (1.50), [NA]	0	0	6 (0.18), [NA]	4 (0.18), [NA]	NR
3 further tests n (%)	1 (0.53), [0]	0	1 (0.05), [NA]	1 (0.07), [NA]	0	0	2 (0.061), [NA]	2 (0.09), [NA]	NR
4 or more further tests n (%)	0	0	0	0	0	0	0	0	NR

FC faecal calprotectin, OR odds ratio, CI confidence interval, IBD inflammatory bowel disease, n number, IBS irritable bowel syndrome, NICE National Institute for Health and Care Excellence, NA not applicable, NR not reported

B15 DEFINING THE APPLICABLE REGION AND SELECTING STUDIES FOR TAILORED META-ANALYSIS

To plot the applicability region I used the mathematical relationship between the test positive rate, the prevalence, the sensitivity and the false positive rate. The mathematical relationship is described in the 2x2 contingency table of test outcome and disease categorisation (Table B25).^{165, 166} As long as a test is better than chance the relationship in equation [1] applies^{165, 166} where s is the sensitivity, f is the false positive rate and r is the test positive rate.

Table B25 Contingency table explaining the mathematic relationship between the test positive rate, the prevalence, the sensitivity and the false positive rate

	With condition	Without condition	Total
Test positive	a	b	a+b
Test negative	c	d	c+d
Total	a+c	b+d	a+b+c+d
	$s=a/a+c$	$f=b/b+d$	$p=a+c/a+b+c+d$ $r=a+b/a+b+c+d$

a true positive, b false positive, c false negative, d true negative, f false positive rate, p prevalence, r test positive rate, s sensitivity

$$\text{if } f \leq s, \text{ this implies } f \leq r \text{ and } s \geq r \quad [1]$$

Data on the test positive rate r can be collected from the setting of interest. The uncertainty of r is expressed as the confidence interval: lower confidence limit (lcl) $< r <$ upper confidence limit (ucl). According to equation [1], the sensitivity for that setting should then take values at least as great as the lower confidence limit ($s \geq$ lcl) and the false positive rate should be no larger than the upper confidence limit ($f \leq$ ucl). The values for the sensitivity and false positive rate can be used to define an area in ROC space. This is the applicable region constrained by the test positive rate, which is depicted in Figure B8 (dashed lines). The applicable region can be further constrained when taking the prevalence p into consideration using the relationship described in formula [2].^{165, 166} The additional constraint reduces the applicable region to a trapezium shape in the ROC space (Figure B8).

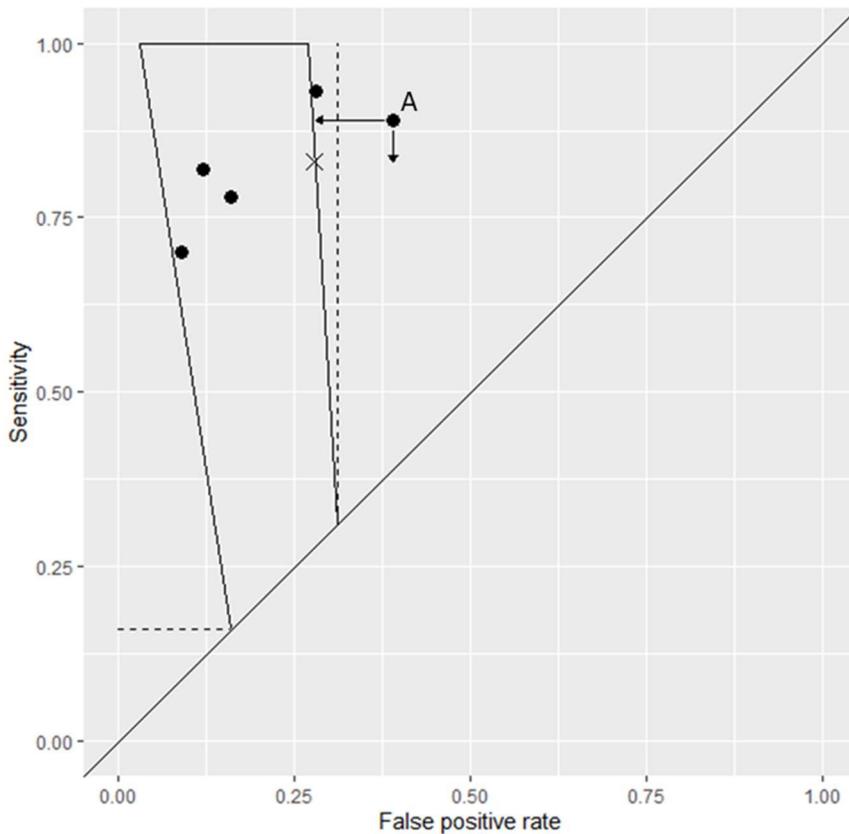


Figure B8 Hypothetical example explaining the applicable region in ROC space (dashed line – constrained by test positive rate only, bold line – constrained by test positive rate and prevalence, circles denote hypothetical study estimates of sensitivity and specificity, x denotes the parameter mu for sensitivity and specificity for study A identified through maximum likelihood estimation.

$$s = \frac{r}{p} - \frac{(1-p)f}{p} \quad [2]$$

Studies with test accuracy estimates lying within this region were considered applicable for the primary care setting. Studies with estimates lying outside the applicable region were assessed for the feasibility of their true population parameters to lie within the applicability region following the methods described by Willis et al. (2014).¹⁶⁵

For those studies with point estimates outside the applicable region (study A in Figure B8), the true population parameter could theoretically lie within the applicable region. This would imply that the study could feasibly lie within the applicable region. The population parameter would need to lie on the boundary of the applicable region (closest to study A) for the greatest probability that the study could feasibly lie within the region. The point closest to the study estimate was the maximum

likelihood estimate after testing 1,000 points along the boundary. The point with the maximum likelihood was taken as the population parameter (Figure B8). The probability was determined that the study's sample statistic is an estimate of the study population with the parameter μ . In cases where the probability was smaller than 0.025 the study was rejected.

B16 R CODE TO SELECT STUDIES FOR TAILORED META-ANALYSIS

```
#####
#
#   SELECTING STUDIES WHICH
#   LIE IN THE APPLICABLE REGION
#   FOR DTA META-ANALYSES
#
#   Author: Brian H Willis (adapted for my tailored meta-analysis of FC testing for IBD)
#####

#####
# Output is Study
#####

#####
# Input values - need to be inputted manually
#####
#IBD versus non-IBD threshold 50, 6 months follow-up to identify IBD diagnoses;
#ICI and uCI of test positive rate and prevalence
r1=0.378415019
r2=0.430840862
p1=0.026551188
p2=0.046468417

z=1.96 # Set levels for confidence intervals in normal distribution

alpha=0.025 # Set levels of significance for binomial probabilities

n=14 # Number of studies in systematic review

#####
# Reads file located on desktop

# File has to have following format: ID Year tp fp fn tn
# Not lower case for true positives, false positives etc

setwd("H:/Documents/PhD/Tailored meta-analysis")
File_Name=read.csv("H:/Documents/PhD/Tailored meta-analysis/use_2x2 tables ibd vs non-ibd
threshold 50_14 studies.csv")

n = nrow(File_Name)

#####
# Define variables

Name=File_Name$Study_ID
Year=File_Name$Year
nf=File_Name$fp+File_Name$tn
```

```

ns=File_Name$tp+File_Name$fn
sens=File_Name$tp/ns
fpr=File_Name$fp/nf
selectBin<-""
selectBin2<-""
flc<-0
fuc<-0
slc<-0
suc<-0

tp<-File_Name$tp
fp<-File_Name$fp
fn<-File_Name$fn
tn<-File_Name$tn

#####
#
# Using Knowledge of test positive rate only
#
#####
# Calculate binomial probability of study estimate being fpr
#or more extreme given parameter lies on the boundary
#####

probab=pbinom(File_Name$fp-1,nf,r2)
signif=1-probab

#####

for (k in 1:n)
{
  if (signif[k]<alpha) selectBin[k]="Reject B1"
  else selectBin[k]="Include B1"

  if (fpr[k]>sens[k])
  {
    signif[k]=555
    selectBin[k] ="Reject B1"
  }
}

selected <- data.frame(Name,Year,tp,fp,fn,tn,sens,fpr,signif,selectBin)

#####
#
# Using Knowledge of test positive rate and Prevalence
#
#####

# Number of points on boundary sampled to calculate maximum likelihood estimate
sample_points=1000

cum_probf<-0
cum_probs<-0
F_MaxLik<-0
S_MaxLik<-0

```

```

ApplicProb<-0

for (k in 1:n)
{
  ApplicProb[k]=0
  cum_prob[k]=0
  cum_probs[k]=0
  F_MaxLik[k]=0
  S_MaxLik[k]=0
}
reject_count =0
i=1
j=1

# Dummy loop counters to check if loops are being executed.
loop1=0
loop2=0
loop3=0

while (i<n+1)
{
  j=i
  # first identify whether sens is greater/equal to fpr if not remove
  if (fpr[i]>sens[i])
  {
    loop1=loop1+1
    reject_count = reject_count + 1
    ApplicProb[i]=555
    selectBin2[i] ="Reject B2"
    j=j+1
  }

#####
#
# Compare study false positive rate with false positive rate on boundary
# for the same sensitivity to see if outside RIGHTMOST boundary
#
#####

# First test study is above F=S line discard if not.
if (sens[i]>=fpr[i])
{

  # For the study sensitivity calculate the corresponding false positive rate on the boundary right most
  boundary
  s=sens[i]
  f=(r2/(1-p1))-(p1*s/(1-p1))

  # If study fpr is to the right of boundary continue, otherwise
  if (fpr[i]>f)
  {
    loop2=loop2+1

  # Test where intercept of outer boundary lies
  sdummy =(r2/p1)
  if (sdummy <=1) s_start=sdummy

```

```

else s_start=0.9999          # R binom cannot deal with proportions above 1
fstart = (r2/(1-p1))-(p1*s_start/(1-p1))
fstop=r2
diff=fstop-fstart
incred=diff/sample_points

# Set up starting values for f, probaility of f, probability of s
f=fstart
FalPosRate=0
TruPosRate=0
Likelih=0
MaxLike =0
probF=0
probS=0
fpr_bound=0

# Any point on boundary is the potential parameter for f and s. Therefore, consider potential
parameters along whole of boundary.
for (count in 1:sample_points+1)
{
  s=(r2/p1)-(1-p1)*f/p1
  if (s>=1) s=1
  if (s<=0) s=0

  # Find maximum likelihood estimate subject to boundary constraint
  probF[count]=dbinom(File_Name$fp[i], nf[i],f)
  probS[count]=dbinom(File_Name$tp[i], ns[i],s)
  FalPosRate[count]=f
  TruPosRate[count]=s
  Likelih[count]=probF[count]*probS[count]

  if(Likelih[count]>MaxLike)
  {
    MaxLike=Likelih[count]
    F_MaxLik[i]=f
    S_MaxLik[i]=s
  }
  f=f+incred
}

# Determine probability of study sens & fpr being more extreme than ML estimate.

cum_probF[i]= 1-pbinom(File_Name$fp[i],nf[i],F_MaxLik[i])
cum_probs[i]= 1-pbinom(File_Name$tp[i],ns[i],S_MaxLik[i])
ApplicProb[i] =cum_probs[i]*cum_probF[i]

if (ApplicProb[i] <alpha)
{
  selectBin2[i]="Reject B2"
  reject_count=reject_count+1
}
else selectBin2[i]="Include B2"
j=j+1
}
}
}

```

```
#####
```

```

#
# Compare study false positive rate with false positive rate on boundary
# for the same sensitivity to see if outside LEFTMOST boundary
#
#####

if (sens[i]>=fpr[i])
{
  s=sens[i]
  f=(r1/(1-p2))-(p2*s/(1-p2))
  if (fpr[i]<f)
  {
    loop3=loop3+1

    # Test where intercept of outer boundary lies
    sdummy =(r1/p2)
    if (sdummy <=1) fstart=0
    else fstart=(r1/(1-p2))-(p2/(1-p2))      # set f_start to where sdummy=1
    fstop=r1
    diff=fstop-fstart
    increm=diff/sample_points

    # Set up starting values for f, probability of f, probability of s
    f=fstart
    FalPosRate=0
    TruPosRate=0
    Likelih=0
    MaxLike =0
    probF=0
    probS=0
    fpr_bound=0

    # Any point on boundary is the potential parameter for f and s. Therefore, consider potential
    parameters along whole of boundary.

    for (count in 1:sample_points+1)
    {
      s=(r1/p2)-(1-p2)*f/p2
      if (s>=1) s=1
      if (s<=0) s=0

      # Find maximum likelihood estimate subject to boundary constraint
      probF[count]=dbinom(File_Name$fp[i], nf[i],f)
      probS[count]=dbinom(File_Name$tp[i], ns[i],s)
      FalPosRate[count]=f
      TruPosRate[count]=s
      Likelih[count]=probF[count]*probS[count]
      if(Likelih[count]>MaxLike)
      {
        MaxLike=Likelih[count]
        F_MaxLik[i]=f
        S_MaxLik[i]=s
      }
      f=f+increm
    }

    # Determine probability of study sens & fpr being more extreme than ML estimate.
    cum_probF[i]= pbinom(File_Name$fp[i],nf[i],F_MaxLik[i])

```

```

cum_probs[i]= pbinom(File_Name$tp[i],ns[i],S_MaxLik[i])
ApplicProb[i] =cum_probs[i]*cum_probf[i]

if (ApplicProb[i] <alpha)
{
  selectBin2[i]="Reject B2"
  reject_count=reject_count+1
}
else selectBin2[i]="Include B2"
j=j+1
}
}

# Remaining studies must be contained within applicable region
if (j==i)
{
  ApplicProb[i]=99
  selectBin2[j] ="Include B2"
}

i=i+1
}

Study<-data.frame(selected, ApplicProb,selectBin2)

write.table(Study,file="H:/Documents/PhD/Tailored meta-
analysis/FileResults_IBD_vs_nonIBD_threshold_50.csv", sep = ",", col.names = NA, qmethod =
"double")

#-----

#Meta-analysis of selected studies

#-----

```

Appendix C Influences on GPs' referral decision and the effects of FC testing

C1 INDICATIVE INTERVIEW SCHEDULE

Introduction:

Thank the interviewee for agreeing to the interview and check if it is OK to record the interview. [Turn on digital voice recorder] Give a brief introduction of yourself and what the aim of the interview is. Ask the interviewee if they have any questions before the interview begins.

Say that you are here to talk to them about your study about GPs' views and experiences related to referral of patients with suspected inflammatory bowel disease or irritable bowel syndrome and ask them to focus on this patient group during the interview.

Topic guide and indicative interview schedule:

Experience with / perception of IBD/IBS patients (including heartsink patients)

- What goes through your mind when a patient presents with typical symptoms for IBS / with persistent abdominal problems?
- What do you think are reasons / triggers for the patient to come and see you?
- *How do you think the symptoms (of IBS) impact on your patient's quality of life?*
- *What are the main challenges when dealing with this group of patients?*

GPs' perceived role in diagnosing IBD/IBS

- What do you think your role is in the diagnosis of these patients?
- If and when in your experience is IBD part of the differential diagnosis?
- How much and in what way would a diagnosis help with the presented symptoms?

Reasons for referral

- What sort of things do you take into account when you decide what to do next?
- In what circumstances would you consider referral of IBS patients?
- Why?

GPs perception of uncertainty and risk

- What are possible consequences (positive and negative) of referral for patients?
- How important are these for your decision on whether to refer?
- What uncertainties are you dealing with at that time and how do they affect your decision to refer / not to refer?

Influences on and thinking behind referral decisions

- What are the main influences on the referral decision?
- What do you expect from referral of IBS patients?
- How does that compare to when you refer patients with suspected IBD?
- How do you manage new presentations of IBD?

Role of patients on referral decisions

- Are there times the patient influences your decision to refer / not to refer?
- Can you give me an example?

Test outcomes and referral decisions

- What tests are available to you and when would you use them?
- In what way would a test result influence your referral decision?
- Are there times in which you would not act on the test results?
- What do you need to know about a test before you would consider using it?

Ending the interview:

Ask them if there is something else they would like to add and what main message they would really like you to take away from the interview.

Ask the interviewee whether they have any questions they would like to ask.

Close the interview and thank the interviewee for their participation.

C2 CODING FRAMEWORK FOR INDEXING INTERVIEW TRANSCRIPTS

<p>Relationship with patients</p> <ul style="list-style-type: none"> • Triggers for help seeking • Expectations of help seeking • Impact of symptoms on patients • Patient centeredness • Helping patients to help themselves • Managing (unrealistic) expectations and conflict 	<p>Diagnosis</p> <ul style="list-style-type: none"> • Approaches to diagnosis • Importance (for primary care) • Cues and nous • IBS versus referral • Many differentials 	<p>Referral</p> <ul style="list-style-type: none"> • Expectations from referral • Patient pressure • Consequences of referral • Reasons and timing for referral • Other influences • Wasted referral 	<p>Management of condition</p> <ul style="list-style-type: none"> • Ineffective treatment and no cure to offer • Keeping the door open • Patient pathway • Managing the person • Alternative management
<p>GP characteristics/values</p> <ul style="list-style-type: none"> • Self-assessment • Ambitions / professional pride • Types of patients • Perceived role of GP • “Starting with a blank sheet” 	<p>Risks and Uncertainties</p> <ul style="list-style-type: none"> • “What if...” / have I done everything • Playing safe • Finding a diagnosis (without specific test) 	<p>Testing</p> <ul style="list-style-type: none"> • Purpose of testing • Role / place / time of FC test • When / who to test • Advantages / impact / consequences of (FC) testing in primary care • Requirements for testing • Opinions on FC testing • Management of tested patients / Testing in context 	<p>Healthcare system</p> <ul style="list-style-type: none"> • Working the system • Availability of resources • ‘Carrying the baton in between times’ • Advice and guidance • Relationship with specialist • Relationship with / view of secondary care • Justifying referral • System

C3 CATEGORISATION OF CODED ELEMENTS INTO HIGHER ORDER CATEGORIES AND SUBTHEMES FOR THE DOCTOR-PATIENT INTERACTION THEME

Table C1 Categorisation of coded elements into higher order categories and subthemes for the doctor-patient interaction theme

Elements	Key dimensions	Categorisation	Higher order classification (subtheme)
Reassuring patient without referring (GP05)	Escalation by clinical need	Prevailing clinical need	Professional clout / authoritarian / GP led
GP doesn't give patients everything they want only what they need (GP03)			
Not referring or investigating if it's not indicated or not going to change anything (GP05)			
Testing for reassurance not helpful (GP09)			
GP refer if they need special test for diagnosis (GP10)			
Diverting people away from investigations that are not needed (GP19)			
Referral is clinical decision (GP01)			
GPs don't think about practical things (GP01)			
Guidelines are only a guide (GP14)			
High faecal calprotectin, it was, like, no, you're going (GP04)			
If FC is high it's easy, GP just refers (GP04, GP08)			
GP tries sticking to guns (GP01)	Sticking to guns		
May end up in a bit of a stand-off (GP03, GP05)			
GP managed to bide a couple of years (GP05)			
Conflict is part of the job (GP03, GP06)			
If explaining doesn't work GP just have to say no (GP03)			
Doesn't mind saying no if it's not needed (GP03)			
Conversations about patient paying taxes but NHS is not means tested (GP03)			
Not being dictated by patient (GP04)			

Different personality might put their foot down (GP06)			
Trying not to over medicalise IBS (GP06)	Unmedicalising IBS	Unmedicalising symptoms	
Symptoms affect patient but it's nothing medical (GP07)			
Don't look for other label (GP08)			
All clinicians can do is say deal with symptoms as best as (GP01)	Handing responsibility to patient		
Try and adapt and live with it (GP01)			
Patients need to take responsibility for actions and bowel movements (GP04)			
Need to learn how to manage on a daily basis (GP04, GP07)			
Asking patients to go and try things (GP04, GP07)			
Up to patient to work out what works for them (GP04, GP05)			
Need to take ownership (GP05)			
Patients need to understand their disease (GP05)			
They need to live with condition (GP06)			
Need to understand what it is and it is nothing serious (GP08)			
Encouraging patients to get medications over the counter (GP15)			
Work their way around it (GP08)			
Giving patients tools (GP05)	Enabling patients	Empowerment	Balance of interests (equipoise)
Make them realise that emotional state has play on symptoms (GP05)			
Make them realise it may be with them for life (GP05)			
Trying to empower patients to manage autonomously (GP06)			
Getting people on board and help them to move on (GP17)			
Getting people to try and belief in medication and GP's advice (GP19)	Encouraging patients		
Providing all diet and lifestyle advice (GP07)	Providing information		

Say to people there is a physical and psychological link (GP17)			
Trying to help patients to manage symptoms (GP06)	Support		
Manage condition themselves alongside their doctor (GP06)			
Looking together at life and what they have tried (GP07)			
Trying to help patient as a whole (GP12)			
Trying to help with anxiety and depression (GP12)			
Helping patients to understand triggers (GP15)			
Working together as a team (GP05, GP12)	Joint working	Shared responsibility	
Decide together on a management plan (GP19)			
Management as a conversation not a didactic thing (GP05)			
Referral with patient's consent (GP01, GP02, GP05, GP12)			
Asking for opinion and what to do (GP04)			
Respecting patient wishes (GP04)			
Happy to give GP time to come to a diagnosis (GP07)			
Choice to refer or keeping it up the sleeve (GP02)	Patient choice		
Explanation that test won't find anything but patient can have it for reassurance (GP17)			
Informing patients fully but guidelines are only guidelines (GP14)			
Never say patient can't have something (GP17)			
GP doesn't refer if patient doesn't want to (GP14)			
Respecting patient's wishes (GP12)			
Patients choose where to be referred to (GP08)			
Patients shop around for GPs (GP07)			
Giving elderly patient the option of referral (GP07)			
Refusal of investigation is patient's choice (GP04)			
Patient might choose to decline (GP02)			
Sit with patients and explain results (GP19)	Importance of communication		
Conversation about specialist tests (GP17)			
Can't treat patients like algorithms, GP needs to listen (GP17)			
Explanation of possibility of IBD before referral (GP13)			

Conversation about role of test (GP09)			
Confidence in explanation more important than explanation itself (GP09)			
Communicating when GPs refer and when they wait for things to develop (GP07)			
Explaining when seeing a specialist won't change anything (GP05)			
Discussing inappropriateness of referral of end of life patient with patient and relatives (GP03)			
Management depends on patient, their worries and symptoms (GP13)	Individualised / tailored care		
Tailoring discussion to patient (GP05)			
GP rolls with it (GP09)	Patient focus		
Consultation is patient focused and patient led (GP07)			
Patients need to be confident to talk (GP07)			
Building relationship with patient (GP12)			
GP can't just assume (GP06)			
GP can offer patient to see a colleague (GP05)	Finding some middle ground	Responsiveness	
Negotiating: refer and try medication whilst waiting (GP06)			
Considering things primary care can do (GP09)			
Explaining what secondary care would do and offering it in primary care (GP09)			
Needing to be diplomatic (GP04, GP06)	Professionalism		
Working it through as a negotiation (GP09, GP14)			
They will come round (GP04)			
It's about working with the patient (GP05, GP07, GP14)			
Knowing the patient (GP05)			
Managing patient long term (GP06)			
Relationship is important (GP06, GP12)			
Finding ways that are not confrontational (GP09)			
Asking what referral is for (GP09)			
Complex procedure (GP14)			
Being honest with patient (GP14)			
Avoid confrontation (GP06)			

Keeping the patient on board (GP06)			
GP has to give them "a fair go" (GP03)			
Relationship with patient influences decisions (GP01)	Counselling		
Spending more time with patient can help (GP12, GP18)			
Listening to patients (GP12)			
Letting them talk to get rid of anxiety (GP14)			
Showing their working gets patients on side (GP14)			
IBS is a time consuming appointment (GP12)			
Spend time with patient to talk about management (GP12)			
GP will never withhold a referral (GP17)	Respect for patient's preferences		
GP is patient's advocate if they want referral (GP17)			
GP tends to follow request (GP11)			
What they want is what they get (GP08)			
Not just about saving money (GP10)			
Establish patient's expectations (GP06, GP08, GP09, GP12, GP15, GP17)	Understanding why patient is seeking help		
Unpicking the underlying reason of worry (GP04, GP06, GP07, GP12, GP15)			
GP needs to read patient (GP09)			
GP training ICE: ideas, concerns, expectations (GP17)			
Advice on diet (GP01, GP09)			
For investigations (GP01, GP19)			
Something to ease their symptoms / symptom relief (GP03, GP05, GP06, GP09, GP10, GP11, GP12, GP14, GP15, GP19)			
Reassurance (GP05, GP07, GP13, GP19)			
Some will want a diagnosis / find out what's wrong (GP05, GP07, GP08, GP09, GP10, GP11, GP12, GP14, GP15)			
Often they just want a label (GP14)			
An explanation and a reason for problems (GP15)			
A referral for a specialist opinion (GP19)			

Worry about a serious pathology they have read / heard about (GP01, GP03, GP04, GP05, GP07, GP08, GP10, GP11, GP12, GP13, GP19)			
Concerned about a nebulous thing (GP01)			
Worried about cancer (GP01, GP03, GP05, GP07, GP10, GP13, GP15, GP17, GP19)			
Abdominal pain a secondary reason "and also I have got ongoing" (GP03, GP14)			
Because symptoms interrupt their life (GP04, GP05, GP06, GP08, GP12, GP14)			
Because someone else has told them they should see a doctor (GP04, GP08, GP11, GP14)			
Ongoing symptoms (GP05, GP06, GP08)			
When over the counter medicine is not effective (GP06)			
More unsure how to handle abdominal pain, lower threshold to see GP (GP06)			
Something new or a change in symptoms (GP08, GP11, GP12, GP18)			
To get rid of the symptoms and pain (GP10)			
Patients can get a bit cross (GP04)	Understanding the patient (needs)		
They need reassuring and explaining (GP04)			
Patients need explanation (GP07)			
Patients need to be taken seriously (GP07)			
Understanding patient belief, expectations and worry (GP07)			
Some people are quite disabled by IBS (GP01, GP08, GP11)			
Restricting people's lifestyles (socially inhibiting), diet and work (GP01, GP02, GP03, GP04, GP05, GP06, GP07, GP08, GP10, GP13)			
Can be embarrassing (GP02, GP04, GP06)			
Symptoms cause concern / impact on emotional health (GP05)			

Uncertainty about future with IBD (GP08)			
Impact on child bearing and contraception with IBD (GP08)			
Symptoms can be trivial or all-consuming or anything in between (GP09, GP11, GP13)			
Avoid referral because of hassle for patients (GP01, GP04, GP14)	Logistics for patients		
Logistics not being able to afford taxi (GP17)			
Patients dislike paying for parking (GP17)			
Patient less in control of appointments (GP06)			
Time away from work for patients (GP06)			
Interruptions to patient's life (GP06)			
GP's perception of patient's anxiety / state of mind (GP1, GP4, GP6, GP10, GP15)	Empathy with patient	Indirect patient pressure	Patient pressure
GP is empathetic (GP1)			
GP feels for the patient (GP07)			
GP feels sorry for patient (GP04)			
Low threshold for helping (GP04)			
Pressure to get to the bottom of the symptoms (GP13)	End of the road		
GP feels he has no alternative (GP5, GP6, GP8)			
Needing a second opinion (GP8)			
GP turning to specialist for help (GP5)			
It reached a point (GP11, GP13)			
GP can't find anything wrong (GP14)			
As a consequence of how consultation has gone over period of months (GP7)			
When GP got to the end of the ladder / road and nothing has helped (GP03, GP07, GP08, GP12, GP13, GP14, GP15, GP16, GP18)			
If GP "is not winning" / is stuck (GP07, GP12, GP16)			
If GP doesn't have a clue what's wrong (GP04, GP05, GP07, GP14)			
Nothing works and patient is getting worse stressful for GP (GP04)			
GP is a bit wavy, patient has lost confidence in GP (GP02)			

Lack of skills, tools or ability to diagnose or instigate management (GP05, GP08, GP14, GP15, GP17,GP19)	Limitations of primary care		
If symptoms so complicated that GP is unsure (GP01, GP18)	Limitations of competence		
Symptoms reach the GP's (individual) threshold for referral (GP11, GP16)			
Unusual or "interesting" patients (GP16)			
Patients vote with their feet (GP4, GP5, GP6, GP7, GP8, GP10, GP11, GP13, GP14, GP16, GP18)	Badgering patient		
Don't want to look at their diet (GP03)	Uncooperative patient		
GP can be falsely reassured by patient and miss really serious things (GP02)			
Irritating if patients choose to overrule an indication (GP09)			
It's a constant battle (GP14)			
People want a magic cure (GP03, GP04, GP07, GP10)	Unrealistic expectations		
Migrants often expect what they had back home (GP04)			
Patients want a definitive answer (GP08)			
Expecting instant diagnosis (GP10)			
GPs get worn down (GP6)	Giving in		
GP feels pressurised (GP6)			
GP being left in difficult position (GP5)			
GP having to refer (GP5)			
GP feels their arm is twisted (GP19)			
GP feels duty bound to refer (GP5)			
Occasionally giving in (GP01)			
Giving in if threatened (GP05)			
GP feels obliged to investigate (GP10)			
Preferring referral to broken down relationship (GP06)			
Like antibiotics for sore throats (GP01)			
Completely irrational decision (GP01)			
Patient cannot be reassured (GP06)	Lack of / losing trust in GP	Direct patient pressure	
Patient fails to see that referral will not change anything (GP05)			
Dissatisfied patient (GP08)			

Patient loses confidence in management (GP02, GP05, GP13, GP19)			
Lacking rapport (GP08)			
Patient demanding a more rapid response than GP can offer (GP06)			
Patient not satisfied without test (GP01, GP05)			
Wondering how GP could know without test (GP01)			
Needing almost objective evidence (GP05)			
Surgery tests and symptom relief not enough (GP05)			
Patient always worries despite GP's reassurance (GP05)			
Patient not happy despite advice want referral (GP10)			
Chatting to GP doesn't satisfy patient (GP10)			
Don't rely on GP's expertise and some blood tests (GP10)			
Patient is not happy with reassurance (GP02, GP13, GP16)			
Patient does not accept the diagnosis (GP08, GP10)			
Worried / struggling patients not coming to terms (GP05, GP06, GP13, GP14)			
When patient is not going to accept what GP is saying (GP05)			
Unable to match up to the patient's expectations (GP08)			
At the point when actual management relationship has broken down (GP06)	Loss of relationship		
Frustrated patients who have already seen other GPs (GP18, GP19)	Reaching for straws		
Patient was keen to be seen by specialist (GP11)			
Mutual decision (GP11)			
If the patient asks for it (GP08, GP11, GP19)			
If patient where nothing has worked press for a referral (GP07)			
Patient demanding referral about what they think they have (GP14, GP15, GP19)	GP as an instrument into secondary care		
Patients having been told that they need referring / investigating (GP18)			

Patient's agenda to see specialist (GP7, GP10, GP11, GP13, GP17, GP18, GP19)			
Referral is goal of consultation (GP03)			
GP feels like secretary writing referral letters (GP09)			
Vocal / assertive / aggressive patient (GP5, GP17)	Confrontational behaviour		
Patients with additional emotional issues (GP5)			
Patients rant and rave (GP5)			
Patient not willing to try management (GP3, GP6)			
Population feels scans are the answer (GP10)	Expectation of investigation in population	Societal expectation	Societal pressure
People want scans (GP10)			
Scans are the end to all (GP10)			
Expectation in population something must be done (GP10)			
People have high expectation of imaging (GP19)			
Patients less tolerant of uncertainty and claim quickly (GP11)	Intolerance of uncertainty and mistakes	Societal intolerance	
Referral is safety netting to get the diagnosis correct (GP11)			
Looking back and pointing finger changed GP's practice (GP11)			
Risk for the GP in terms of litigation if GP is wrong (GP01)	Medico-legal pressure		
Guidelines for everything so that GPs can defend themselves by saying referral was according to guidelines (GP03)			
Experience with legal action makes GPs more defensive in their medicine and refer more (GP03)			
If GP refers everybody no one would be taking legal action because GP hasn't missed anything (GP03)			
In this medical legal world GPs can't deny patient referral (GP05)			
Defensive practice of GPs who don't like risk element (GP05)			
Everybody and his dog thinks they can go to the GMC and complain (GP05)			

Medico-legally there is no reward for saving the NHS money (GP11)			
NHS doesn't back up GP, it's safer to refer (GP11)			
In medico-legal practice a guideline will be used as a standard against the GP (GP11)			
Referrals go up because people practicing more medico-legal protection (GP18)			