Non-invasive testing for the diagnosis and assessment of colorectal cancer and adenomatous polyps

Dr Georgia E Woodfield

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

2021

A thesis submitted to Imperial College London for the degree of
Doctorate of Philosophy

Supervisors

Professor George B Hanna
Department of Surgery and Cancer, Imperial College London
St Mary's Hospital, London UK

Professor Amanda J Cross
Department of Surgery and Cancer, Imperial College London
St Mary's Hospital, London UK

Dr Ilaria Belluomo
Department of Surgery and Cancer, Imperial College London
St Mary's Hospital, London UK

STATEMENT OF ORIGINALITY

I declare that I am the sole author of this thesis and all work within it is my own. Individuals who collaborated in any part of this work are appropriately acknowledged.

Signed: Georgia Woodfield Date: 27th April 2021

G E Woodfield

COPYRIGHT DECLARATION

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

ABSTRACT

Colorectal cancer (CRC) is the 2nd most common cause of UK cancer death and makes up 11% of new cancer cases. However, 5-year survival rates are 58%, amongst the lowest in Europe. There is need for improvement because, if detected early, CRC survival rates can exceed 90%. A test measuring breath volatile organic compounds (VOCs) could provide an attractive non-invasive tool for early CRC detection, aiming to augment and streamline existing referral pathways, and ultimately save lives.

The first aim of this PhD was methodological development aiming to improve reliability and standardisation of measurement and interpretation of breath VOCs. This culminated in the development of a quality control system, providing a framework for detection of inadequate quality breath samples prior to analysis. The second aim was to determine the diagnostic accuracy of breath for CRC and adenomatous polyp diagnosis, using the ReCIVA® breath sampler and thermal desorption tubes. 1432 patients having a colonoscopy or colorectal adenocarcinoma resection over 7 London hospitals had their breath analysed using gas chromatography mass spectrometry (GC-MS). Machine learning algorithms showed that CRC patients (n=162) could be predicted from positive and negative controls (n=1270) with a sensitivity of 77%, specificity of 87%, and an area under the Receiver Operating Characteristic (AUROC) curve of 0.87. Fifteen of the most predictive VOCs were from sulphur containing, alcohol, alkane, ester or phenol chemical groups. Colorectal polyp patients (n=592) could be predicted from positive and negative controls (n=619): sensitivity 66%, specificity of 54%, AUROC curve of 0.65. The third aim was to demonstrate feasibility and acceptability of breath testing in a large primary care study.

These exciting findings represent advancement in knowledge of diagnostic CRC VOCs, paving the way for larger targeted studies to validate findings, establish quantitative thresholds and move towards eventual implementation of breath testing into clinical practice.

TABLE OF CONTENTS

Statement of originality 2

Copyright declaration 2

Abstract 3

Table of contents 4

List of figures 10

List of tables 13

Abbreviations 15

Peer reviewed publications 18

Acknowledgments 19

1 Chapter 1: Introduction 22

1.1 Colorectal cancer and adenomatous polyps 23

1.1.1 Diagnosis of colorectal cancer and colorectal polyps 24

1.1.2 Barriers to early diagnosis 29

1.1.3 Summary: the need for new biomarkers for colorectal cancer 30

1.2 Biomarkers for colorectal cancer and adenomatous polyps 31

1.3 Volatile organic compounds 34

1.3.1 Definition and background 34

1.3.2 Mechanism of production 35

1.3.3 Volatile organic compounds in current clinical practice 37

1.3.4 Breath testing for cancer using volatile organic compounds 37

1.4 Rationale and aim of PhD project 38

2 Chapter 2: Systematic literature review of volatile organic compounds in colorectal disease 40

2.1 Methods 41

2.1.1 Search strategy and selection criteria 41

2.1.2 Quality Control 42

2.2 Results 43
2.2.1 Benign diseases of the colon 44
2.2.2 High-risk colorectal polyps 45
2.2.3 Colorectal cancer 45
2.3 Discussion 54
2.4 Conclusion 56
3 Chapter 3: methods and measurement of breath collection, and ensuring robust quality control 57
3.1 Breath collection systems 58
3.1.1 Online breath sampling 58
3.1.2 Breath bags 59
3.1.3 Thermal desorption tubes 60
3.2 ReCIVA® breath collection system 61
3.3 Measurement of volatile organic compounds using mass spectrometry 64
3.3.1 Breath analysis using Thermal Desorption Gas Chromatography Mass Spectrometry 65
3.3.2 Breath analysis using Selective Ion Flow Tube Mass Spectrometry 66
3.3.3 Breath analysis using Proton-Transfer-Reaction Time-of-Flight Mass Spectrometry 66
3.3.4 Sensor technology 67
3.4 Current challenges in measuring exhaled breath 67
3.4.1 Sampling breath using the ReCIVA® breath sampler and thermal desorption tubes 68
3.4.2 Complexity of breath as a matrix 69
3.4.3 Variation and lack of standardisation of analytical methods 69
3.5 Development of a quality control process for breath testing 70
3.5.1 Thermal desorption tubes quality control 70
3.5.2 Breath sampling equipment quality control 72
3.5.3 Retrospective ReCIVA® trace quality control checks 74
3.5.4 Lab instrument quality control checks 76
3.6 Creation of a sample quality control system (breath content quality control), for thermal desorption tubes 77
3.6.1 Rationale 77
3.6.2 Aim 77
3.6.3 Methods 78
3.6.4 Results 80
3.6.5 Discussion 82

3.6.6 Conclusion 83

3.7 Summary of chapter 83

4 Chapter 4: Profiling of volatile organic compounds in breath for the detection of colorectal cancer and adenomatous polyps, the Colorectal Breath Analysis (COBRA) study 84

4.1 Study Aim 85

4.1.1 Objectives 85

4.2 Ethics 85

4.3 Methodology 85

4.3.1 Inclusion Criteria: 86

4.3.2 Exclusion Criteria: 86

4.3.3 Patient selection - endoscopy unit 86

4.3.4 Patient selection - theatres 87

4.3.5 Target study size calculation 87

4.3.6 Patient sampling 88

4.3.7 Breath sample handling 89

4.3.8 Specimen analysis 89

4.3.9 Rationale for measurement using different analytical platforms 89

4.3.10 Staff and Training 90

4.3.11 Quality Control 91

4.3.12 Data analysis 91

4.4 Results 96

4.4.1 Recruitment 96

4.4.2 Clinical details and demographics of recruited patients 100

4.4.3 Clinical details of colorectal cancer patients 106

4.4.4 Sample processing times 110

4.4.5 Results of initial univariate statistics 110

4.4.6 Results of machine learning prediction model, analysis 1: colorectal cancer vs. controls 111

4.4.7 Results of machine learning prediction model, analysis 2: colorectal cancer vs. no colorectal pathology 120

4.4.8 The esters 120

4.4.9 Sulphur compounds 124

4.4.10 The alkanes 125

4.4.11 The alcohols 128

4.4.12 Phenol 132

4.4.13 The non-aromatic cyclic hydrocarbons 133
4.4.14 Exploration of potential patient-related confounders 135
4.4.15 Results of machine learning prediction model, analysis 3 - colorectal polyps 140
4.5 Discussion 144
4.5.1 The esters 145
4.5.2 Sulphur compounds 146
4.5.3 The alkanes 147
4.5.4 The alcohols 149
4.5.5 Phenol 150
4.5.6 The non-aromatic cyclic hydrocarbons 151
4.5.7 Potential significance of the COBRA findings 151
4.5.8 Discussion of potential patient related confounders 153
4.5.9 Sampling considerations and limitations 158
4.5.10 Gas chromatography-mass spectrometry considerations and limitations 159
4.5.11 Gas chromatography mass spectrometry column choice 160
4.5.12 Study design considerations and limitations 160
4.5.13 Limitations caused by data loss and exclusion 163
4.5.14 Human factors 163
4.5.15 Establishing breath analysis as a diagnostic or monitoring tool 163
4.5.16 Future work 165
4.6 Conclusion 166
4.7 Funding 167
5 Chapter 5: The Methodological Approaches towards a Gastrointestinal Cancer (MAGIC) breath testing study, based in primary care 168
5.1 Background 169
5.2 Aims of study 170
5.3 Ethical considerations 170
5.4 Methodology 170
5.4.1 Eligibility 171
5.4.2 Breath Sampling 171
5.4.3 Sample processing and breath analysis 172
5.4.4 Training 173
5.4.5 Study recruitment phases 174
5.4.6 Study recruitment strategies 174

5.4.7 Sampling models....177
5.4.8 Assessment of feasibility and acceptability of breath testing....178
5.5 Results....179
5.5.1 Theme 1: Recruitment....180
5.5.2 Theme 2: Sample Collection....191
5.5.3 Theme 3: Storage and transport....196
5.5.4 Theme 4: Breath analysis....197
5.6 Discussion and limitations....198
5.7 Conclusion....203
5.8 Funding....203
6 Chapter 6: Summary, conclusions and future work....204
6.1 Summary....205
6.2 Future work....207
6.3 Conclusion....208
7 Appendix....210
7.1 Appendix 1: Modification of QUADAS-2 assessment tool....211
7.2 Appendix 2: Modified QUADAS-2 assessment breakdown for each included study....213
7.3 Appendix 3: Consent form- Non-invasive testing for gastrointestinal disease....214
7.4 Appendix 4: Participant Information Leaflet - Non-invasive testing for gastrointestinal disease 215
7.5 Appendix 5: Participant Information Leaflet- Non-invasive testing for colorectal disease 219
7.6 Appendix 6: Consent form- Non-invasive testing for colorectal disease....224
7.7 Appendix 7: Baseline clinical data form....225
7.8 Appendix 8: ReCIVA® List of tasks....227
7.9 Appendix 9: COBRA protocol....237
7.10 Appendix 10: Mass spectra and NIST matches for the esters....255
7.11 Appendix 11: Mass spectra and NIST matches for the sulphur compounds....258
7.12 Appendix 12: Mass spectra and NIST matches for the alkanes....259

7.13 Appendix 13: Mass spectra and NIST matches for the alcohols 261

7.14 Appendix 14: Mass spectra and NIST matches for phenol 265

7.15 Appendix 15: Mass spectra and NIST matches for the non-aromatic cyclic hydrocarbons 266

7.16 Appendix 16: Names of 26 GP practices participating in the MAGIC breath test study269

7.17 Appendix 17: Summary of National Institute for Health and Care Excellence guidelines for gastrointestinal cancer referral 2016 270

7.18 Appendix 18: Poster for GPs 271

7.19 Appendix 19: Letter for GPs 272

7.20 Appendix 20: Participant Information Leaflet- MAGIC study 273

7.21 Appendix 21: Consent form – MAGIC study 277

7.22 Appendix 22: Patient acceptability questionnaire 278

7.23 Appendix 23: Questionnaire for GPs 280

7.24 Appendix 24: Breath content quality control for samples run on the PTR-ToF-MS.....282

7.25 Appendix 25: Example of GP database searches 283

7.26 Appendix 26: Optimising recruitment in primary care studies, lessons learnt from the MAGIC breath testing study. 288

8 References and bibliography 289

LIST OF FIGURES

Figure 2-1 PRISMA flow diagram of literature search 43
Figure 3-1 Online breath sampling into the SIFT-MS 59
Figure 3-2 Breath sampling into a Nalophan® bag 60
Figure 3-3 ReCIVA® breath sampler being used and diagram, adapted from the Owlstone ReCIVA® manual 62
Figure 3-4 Laptop screen depicting the breath carbon dioxide and mask pressure traces during sampling 63
Figure 3-5 The ReCIVA® kit packed into a travel-size suitcase 64
Figure 3-6 Workflow of quality control points for breath analysis 70
Figure 3-7 Concentrations of C4 to C10 compounds on 30 cleaned TD tubes, with quality control thresholds indicated. 71
Figure 3-8 CASPER device for filtration of entrained air 72
Figure 3-9 Example of a compliant CASPER VOC trace 73
Figure 3-10 Example of a non-compliant CASPER VOC trace 73
Figure 3-11 Example of the graphically represented output from a correctly performed breath sample performed with ReCIVA® 75
Figure 3-12 Example of the graphical representation of a poorly performed/inadequate breath sample taken with the ReCIVA® 75
Figure 3-13 Comparison of acetone abundance in controls (non-biological) samples vs. human breath samples 81
Figure 4-1 Flowchart outlining the reasons for post sampling data exclusion for the COBRA study, GC-MS dataset 99
Figure 4-2 Frequency of site of 162 colorectal adenocarcinomas 106
Figure 4-3 T stage score distribution in 162 colorectal adenocarcinomas 106
Figure 4-4 N stage score distribution in 162 colorectal adenocarcinomas 107
Figure 4-5 M stage score distribution in 162 colorectal adenocarcinomas 107
Figure 4-6 Frequency distribution of grade of tumour differentiation of 162 colorectal adenocarcinomas 108

Figure 4-7 Distribution of T stage (1-4) at diagnosis of colorectal adenocarcinoma, by route of diagnosis..... 109
Figure 4-8 Age distribution by T stage, 162 colorectal adenocarcinomas..... 109
Figure 4-9 Machine learning analyses for CRC prediction..... 111
Figure 4-10 ROC curve for the prediction of CRC; CRC = 162, non-CRC = 1270..... 112
Figure 4-11 ROC curve illustrating predictive power of 15 VOCs in determining CRC from non-CRC..... 119
Figure 4-12 Boxplot of propyl propionate abundance, VOC 1, in breath in non-CRC vs. CRC. 122
Figure 4-13 Boxplot of allyl acetate abundance, VOC 8, in breath in non-CRC vs. CRC..... 122
Figure 4-14 Boxplot of overlapping ester to allyl acetate abundance, VOC 9, in breath in non-CRC vs. CRC. 123
Figure 4-15 Boxplot of methyl 2-butynoate abundance, VOC 12, in breath in non-CRC vs. CRC. 123
Figure 4-16 Boxplot of dimethyl sulphide abundance, VOC 2, in breath in non- CRC vs. CRC. 125
Figure 4-17 Boxplot of alkane biomarker abundance, VOC 3, in breath in non-CRC vs. CRC. 127
Figure 4-18 Boxplot of alkane biomarker abundance, VOC 11, in breath in non-CRC vs. CRC. 127
Figure 4-19 Boxplot of 3-ethyl-hexane abundance, VOC 15, in breath in non-CRC vs. CRC. 128
Figure 4-20 Boxplot of 1,3-Dioxolane-2-methanol abundance, VOC 4, in breath in non-CRC vs. CRC. 130
Figure 4-21 Boxplot of 2-phenoxy-ethanol abundance, VOC 5, in breath in non-CRC vs. CRC. 130
Figure 4-22 Boxplot of 2,2,4-trimethyl-3-pentanol abundance, VOC 10, in breath in non-CRC vs. CRC. 131
Figure 4-23 Boxplot of 1-undecanol abundance, VOC 13, in breath in non-CRC vs. CRC. 131
Figure 4-24 Boxplot of phenol abundance, VOC 14, in breath in non-CRC vs. CRC. 133
Figure 4-25 Boxplot of cyclopropane abundance, VOC 6, in breath in non-cancer vs. CRC. 134
Figure 4-26 Boxplot of 3,4-dimethyl- 1,5-cyclooctadiene abundance, VOC 7, in breath in non-cancer vs. CRC. 135

Figure 4-27 Median abundance of the 9 VOCs found to be significantly increased in CRC, by T stage at time of breath test, log scale. 138
Figure 4-28 Median abundance of 6 VOCs found to be significantly decreased in CRC, by T stage of CRC at time of breath test, log scale. 139
Figure 4-29 ROC curve of the prediction of normal/benign/IBD patients against all colorectal polyp patients (any polyp severity) 142
Figure 4-30 ROC curve of the prediction of normal/benign/IBD patients against intermediate-/high-risk polyp patients 143
Figure 5-1 Age distribution of all patients with age data (n=980) in the MAGIC study 182
Figure 5-2 Duration of symptoms for 365 patients in Phase-2 of the MAGIC study 182
Figure 5-3 Level of worry for 365 patients in Phase-2 of the MAGIC study 183
Figure 5-4 Patient recruitment percentage by the method used to engage them, Phase-1 185
Figure 5-5 Recruitment for the MAGIC breath testing study 186
Figure 5-6 Influence of age on response to different recruitment strategies (Phase-1) 189
Figure 5-7 Influence of gender on response to different recruitment strategies (Phase-1) 190
Figure 5-8 Influence of ethnicity on response to different recruitment strategies (Phase-1) 191
Figure 7-1 Obtained mass spectrum for VOC 1 compared to the mass spectrum of propyl propionate from the NIST database. 255
Figure 7-2 Obtained mass spectrum for VOC 8 (left) and VOC 9 (right) compared to the mass spectrum of allyl acetate from the NIST database. 256
Figure 7-3 Obtained mass spectrum for VOC 12 compared to the mass spectrum of methyl 2-butynoate from the NIST database. 257
Figure 7-4 Obtained mass spectrum for VOC 2 compared to the mass spectrum of dimethyl sulphide from the NIST database. 258
Figure 7-5 Obtained mass spectra for VOC 3 (left) and VOC 11 (right) branched C14 alkanes. 259
Figure 7-6 Obtained mass spectrum for VOC 15 compared to the mass spectrum of 3-ethyl-hexane from the NIST database. 260
Figure 7-7 Obtained mass spectrum for VOC 4 compared to the mass spectrum of 1,3-dioxolane-2-methanol from the NIST database. 261

Figure 7-8 Obtained mass spectrum for VOC 5 compared to the mass spectrum of 2-phenoxy-ethanol from the NIST database..... 262
Figure 7-9 Obtained mass spectrum for VOC 10 compared to the mass spectrum of 2,2,4-trimethyl-3-pentanol from the NIST database 263
Figure 7-10 Obtained mass spectrum for VOC 13 compared to the mass spectrum of 1-undecanol from the NIST database. 264
Figure 7-11 Obtained mass spectrum for VOC 14 compared to the mass spectrum of phenol from the NIST database 265
Figure 7-12 Obtained mass spectrum for VOC 6 compared to the mass spectrum of cyclopropane from the NIST database. 267
Figure 7-13 Obtained mass spectrum for VOC 7 compared to the mass spectrum of 3,4-dimethyl-1,5-cyclooctadiene from the NIST database. 268

LIST OF TABLES

Table 1-1 Recent studies using proteomic-based approaches to find CRC biomarkers 33
Table 2-1 Search strategy used in Ovid online search engine 41
Table 2-2 Detail of all included studies for VOC testing for colorectal disease using breath, blood, faeces and urine. 47
Table 3-1 Breakdown of samples collected for the breath content quality control study 80
Table 3-2 Acetone abundance ranges for human breath vs. non-biological control samples 82
Table 4-1 Breakdown of COBRA recruitment by NHS centre 96
Table 4-2 Patient approach and decline rate, and pre-test exclusion rate per site for COBRA study 97
Table 4-3 Diagnosis at colonoscopy for 1432 patients 100
Table 4-4 Demographics of included patients, by main pathology groups 103
Table 4-5 Thermal desorption tube storage time (days), n=1432 110
Table 4-6 List of top discriminating features contributing to the differentiation of CRC patients (n=162) from positive and negative control patients (n=1270) 114
Table 4-7 Top 15 cancer-differentiating ions, defined as those with the potential to be CRC biomarkers 118

Table 4-8 Abundance, measured in peak area count, of the four significant esters measured by GC-MS between patients with (n= 162) and without CRC (n=1270) 121
Table 4-9 Abundance, measured in peak area count, for dimethyl sulphide measured by GC-MS between patients with (n= 162) and without CRC (n=1270) 124
Table 4-10 Abundance, measured in peak area count, of three significant alkanes measured by GC-MS between patients with (n= 162) and without CRC (n=1270) 126
Table 4-11 Abundance, measured in peak area count, of four significant alcohols measured by GC-MS between patients with (n= 162) and without CRC (n=1270) 129
Table 4-12 Abundance, measured in peak area count, for phenol measured by GC-MS between patients with (n= 162) and without CRC (n=1270) 132
Table 4-13 Abundance, measured in peak area count, of two significant non-aromatic cyclic hydrocarbons measured by GC-MS between patients with (n= 162) and without CRC (n=1270) 134
Table 4-14 Machine learning cancer prediction statistics for different age brackets, CRC vs. non-CRC patients 136
Table 4-15 Correlation between significant VOC abundances (originally measured in peak area) with fasting time (hours), in 1197 patients with fasting data 137
Table 4-16 Machine learning comparisons, done with different group combinations, to investigate the ability to predict colorectal polyps 141
Table 5-1 Demographics of 998 eligible patients, and characteristics of reported symptoms 181
Table 5-2 Patient enrolment by recruitment method for Phase-1 MAGIC study (633 eligible patients) 184
Table 5-3 Patient sampling models used throughout the MAGIC study 184
Table 5-4 Summary of barriers to recruitment and subsequent implementations, during Phase-1 of the MAGIC breath testing study. 186
Table 5-5 Feedback regarding recruitment strategies from focus group and field notes 188
Table 5-6 Summary of themes regarding limiting factors of the breath sampling process. 192
Table 5-7 Summary of patient acceptability questionnaire responses (n=365, Phase-2 of study) 194
Table 5-8 Pass rate of breath sampling process quality control in Phase-2 of the study 198
Table 7-1 Breakdown of texts sent and responses 287

ABBREVIATIONS

ANOVA: Analysis of variance

AUROC curve: area under the Receiver Operating Characteristic curve

BCSP: Bowel cancer screening programme

BMI: Body mass index

CASPER: Clean Air Supply Pump for ReCIVA®

CCG: Clinical commissioning group

CD: Crohn's disease

CIBH: Change in bowel habit

COBRA: Colorectal Breath Analysis

COX-2: Cyclooxygenase-2

CO_2: Carbon dioxide

CRC: Colorectal cancer

CSOs: Clinical study officers

CT: Computed tomography

DNA: Deoxyribonucleic acid

IBD: Inflammatory bowel disease

IBDu: inflammatory bowel disease unclassified type

IBS: Irritable bowel syndrome

IDA: Iron deficiency anaemia

IMS: Ion Mobility Spectrometry

FAP: Familial adenomatous polyposis

FIT: Faecal immunochemical test

GC-MS: Gas chromatography mass spectrometry

gFOBT: Guaiac faecal occult blood test

GP: General practitioner

GI: Gastrointestinal

GNPS: Global Natural Product Social Molecular Networking

Hb: Haemoglobin

HC: Healthy control

IPAA: Ileal pouch-anal anastomosis

LC-MS: Liquid chromatography mass spectrometry

MAGIC: Methodological Approaches Towards a Gastrointestinal Cancer

MRI: Magnetic resonance imaging

mRNA: messenger ribonucleic acid

NHS: National Health Service

NICE: National Institute of Clinical Excellence

NIHR: National Institute for Health Research

NIST: National Institute of Standards and Technology

NMF: Non-negative matrix factorisation

PDSA cycles: Plan-Do-Study-Act cycles

ppb: Parts per billion

PPIs: Proton pump inhibitors

ppt: Parts per trillion

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

PTR-ToF-MS: Proton Transfer Reaction-Time of Flight- Mass Spectrometer

QC: Quality control

QUADAS-2: Quality Assessment of Diagnostic Accuracy Studies 2

RCT: Randomised controlled trial

REC: Research ethics committee

ReCIVA®: The Respiration Collector for *in-vitro* Analysis

RF: Random forest

ROC: Receiver Operating Characteristic

SCFAs: Short chain fatty acids

SIFT-MS: Selected Ion flow tube mass spectrometry

SOP: Standard operating procedure

SPME: Solid-phase microextraction

STARD: Standards for Reporting of Diagnostic Accuracy Studies

SVM: Support Vector Machines

TD: Thermal desorption

TNM: Tumour, nodes, metastases

UC: Ulcerative colitis

UK: United Kingdom

VOCs: Volatile organic compounds

WP: Work package

2WW: 2 week-wait

PEER REVIEWED PUBLICATIONS

Woodfield G, Belluomo I, Boshier PR, Waller A, Fayyad M, von Wagner C, Cross AJ, Hanna GB. Feasibility and acceptability of breath research in primary care: a prospective, cross-sectional, observational study. *BMJ Open* 2021;**11:**e044691. doi:10.1136/bmjopen-2020-044691

ACKNOWLEDGMENTS

I am greatly indebted to my supervisors **Professor George Hanna, Professor Amanda Cross** and **Dr Ilaria Belluomo** for their continued guidance, support and encouragement, and for always believing in me. I would also like to say a special thanks to **Mr Piers Boshier** for his ever present support, advice and talent for problem solving and to **Professor Patrik Spanel** for his invaluable guidance and generously-given expertise regarding mass spectrometry methods, interpretation of ion chemistry, and for showing me how to be a scientist. I also wish to acknowledge the invaluable input and expertise of **Professor Wendy Atkin**, who was a co-supervisor of this work until she very sadly passed away in December 2018. I must also mention the consistent support and encouragement from my husband, parents, sister and more recently my baby son throughout my PhD journey.

I would also like to specifically acknowledge the help of the following people who contributed to this research.

Dr Ilaria Belluomo – for day to day guidance and support, practical troubleshooting with lab instruments, time spent, and much patient explanation with all lab-related components of the work throughout all of my projects, as well as being an extremely supportive overall supervisor.

Mr Oscar Ayrton- for support and guidance using the laboratory instruments and for prioritising quality control, which greatly benefitted all of the laboratory users on a daily basis.

Mr GengPing Lin - for assistance with sample collection of colorectal cancer patients at St Mary's Hospital. I really appreciate your kindness and team spirit.

Dr Christian Von Wagner, University College London - for guidance regarding the design of the patient acceptability questionnaires for Breath MAGIC.

Dr Kirill Veselkov – for writing the mass spectrometry analysis programme MSHub then for his guidance and expertise in using it for the GC-MS data of this thesis. Also for his input regarding the statistical analysis of the COBRA data using machine learning.

Dr Ivan Laponogov – for his invaluable input and guidance regarding the statistical analysis of the COBRA data using machine learning. Ivan used a high performance computer facility at Imperial College London to run a machine learning pipeline to process all of the abundance data of unidentified features in each patient's breath sample with attached extensive metadata, for the COBRA study (Chapter 4). Thank you also for your hard work and patience in helping me to interpret the results.

Professor Patrik Spanel- for his guidance and advice regarding identity of compounds and interpretation of ion chemistry for the COBRA study.

Dr Andrea Romano- for his guidance and technical support in using the PTR-ToF-MS and helping to develop a method for its use.

Dr Alberto Vidal-Diez – for his guidance and support regarding interpretation of statistics throughout my time at Imperial College.

All the NIHR clinical research network staff, in particular **Miss Lucinda Hetherington**- for helping to coordinate GP surgeries for the MAGIC study at multiple sites over London, and for tirelessly sampling the breath of over 1000 patients.

All the NIHR- affiliated staff working at Charing Cross, St Mary's, St Mark's, Chelsea and Westminster, West Middlesex, St George's and Homerton Hospitals- for sampling breath in endoscopy departments and theatres across London and helping me to coordinate logistics and transport of samples.

Miss Annabelle Waller, Mr Harrypal Panesar and Miss Maya Fayyad –the most engaged medical students I have met in a long time. Thank you for your assistance with data inputting for COBRA and MAGIC.

Mr Ahmed Hosny and Mr Anand Bhundia- GP Network Support Officers at Central London Healthcare Clinical Commissioning Group- for setting up and performing the electronic search of the databases of participating GP practices for identification of potential patients for the hub and spoke model of Phase-2 of the MAGIC study.

Rosetree's and Stoneygate Trusts via **Imperial Health Charity** for supporting the COBRA study via a special purpose fund grant – grant reference 171812a.

NIHR London In Vitro Diagnostic Co-operative programme for providing infrastructure funding for the VOC lab - grant reference P67522.

1 CHAPTER 1: INTRODUCTION

1.1 **Colorectal cancer and adenomatous polyps**

Descriptive statistics

Colorectal cancer (CRC) is the 2nd most common cause of cancer death in the UK, and is the fourth most common cancer in the UK population, making up 11% of UK new cancer cases. (1) The incidence of CRC from the Office for National Statistics cancer registration data in 2017 was 70.6 cases per 100,000 males and 54.8 cases per 100,000 females. (2) The age-standardised 1-year survival rates for men and women who were diagnosed with CRC were 79.3% and 77.1% respectively, diagnosed between 2013 and 2017. 5-year survival rates for men and women diagnosed during the same date range were 58.2% and 58.6%, respectively. (3) Although incidence of CRC remained stable between 2015 – 2017, and over the past decade incidence has decreased by 4%, the survival rates for CRC are amongst the lowest in Europe. (1, 4) The burden of CRC is therefore great, with room for substantial improvement in survival rates.

CRC is an ideal cancer to tackle from a preventative perspective, as it most often follows a slow trajectory of growth from a benign neoplasm (polyp) to cancer. This was first described in 1988; the first stage being the formation of benign polyps, progression to histologically more advanced neoplasms, then the final step of transformation into invasive carcinoma. (5) Polyps for the purposes of this thesis include both adenomatous and serrated polyps. Adenomas are benign glandular tumours of the colorectum, displaying different histopathological appearances varying between predominantly tubular in appearance to predominantly villous or mixed, with low to high grade dysplasia. (6) Tubulovillous/villous lesions and any with high grade dysplasia are more likely to be become malignant. (7) Serrated polyps is an umbrella term used to describe hyperplastic polyps, sessile serrated lesions, traditional serrated adenomas and mixed polyps of the colorectum; all are characterised histologically by a saw- toothed (serrated) appearance of the crypt epithelium, and have different malignant potential depending on size and dysplasia. (7, 8) There are a large number of factors that drive the sequence of changes from polyp to cancer, including the accumulation over time of gene mutations and epigenetic alterations, as well as local inflammatory changes, which themselves may have multiple driving factors. (9) Molecular alterations that have been identified include global DNA alterations such as microsatellite instability and aneuploidy, as well as epigenetic alterations such as CpG Island Methylator

Phenotype (CIMP) (prominent in the progression of sessile serrated polyps to CRC). (9) Specific gene mutation inheritance is also well recognised.

1.1.1 Diagnosis of colorectal cancer and colorectal polyps

Early diagnosis is of great importance in CRC, as it directly affects survival outcomes. If diagnosed early, CRC survival rates exceed 90% (10) and reducing delays to diagnosis is considered key to improving outcomes. (11) The 2020 National Bowel Cancer Audit Report reported CRC diagnosis routes for patients diagnosed in 2018-2019; 54% were diagnosed through referral from their general practitioner (GP) and 10% via the National Health Service (NHS) Bowel Cancer Screening Programme (BCSP). (12) However, a significant proportion, 19%, were diagnosed only after an emergency presentation. (12) There was a direct link between treatment outcomes and route of diagnosis, where 90% of the patients who were diagnosed through screening and 69% of those referred by their GP were able to be treated with curative intent. This contrasted to only 50% of those who presented as an emergency. (12)

The gold standard test for diagnosis of CRC is colonoscopy. This procedure allows a visual examination of the entire length of the large bowel and can incorporate the collection of tissue samples. Any lesions (CRCs or colorectal polyps) that are found have their location recorded and undergo histological diagnosis, which guides decisions regarding surgical/endoscopic resection options, targeted further treatments (e.g. chemotherapy or immunotherapy), as well as need for future colonoscopic surveillance. This makes colonoscopy superior to computed tomography (CT) or magnetic resonance imaging (MRI) scanning in those that are fit enough to have it, for diagnostic purposes. Another benefit is that colonoscopy (and flexible sigmoidoscopy for distal lesions) are also highly sensitive for small lesions, and so neoplasia is detectable at a very early stage. (13) In the case of small lesions, the procedure is both diagnostic and therapeutic as most can be removed in the same visit. This is key for prevention of CRC for those who are found not to have CRC at the time of the test. In some countries, such as the USA and Germany, colonoscopy is a first line test, and is used for wide-scale screening. In Germany the sensitivity and specificity of colonoscopy for identifying polyps and cancers is greater than 98%. (14)

Despite being gold standard for CRC diagnosis, colonoscopy is not used as a first line test either for screening or for all symptomatic patients in the UK for a number of reasons.

Firstly, it would not be feasible in the UK, as the NHS currently has insufficient trained endoscopists to perform the number of procedures that would be required if used first-line for screening or for low-risk symptoms. This is also in the context of ever increasing demand for colonoscopies; Bowel Cancer UK have suggested that there will be a 10–15% increase in demand for colonoscopies every year. (15) The UK currently already performs around 2 million endoscopic procedures per year. (16) The problem of capacity would therefore only worsen over time. Secondly, it may provide limited benefit for patients. In a 2015 review, most colonoscopies performed in symptomatic patients were found not to discover any serious bowel pathology, and most did not result in changes to the treatment given to the patient. (17) Yield would be expected to be even less in a screening population. Thirdly colonoscopy is not without risks, which include abdominal pain, bleeding and in some cases bowel perforation. (18) Serious complications (perforations and major haemorrhage) happen in approximately 3 per 1000 colonoscopies (UK NHS audit data). (19) Fourthly, the extra colonoscopies required would undoubtedly impact negatively on the waiting time for high-risk patients such as those qualifying for the two week wait (2WW) pathway. (20) It has been postulated that risk and cost could be decreased by replacing colonoscopy with flexible sigmoidoscopy in selected patients (those with rectal bleeding or change in bowel habit without anaemia or an abdominal mass), because they are very unlikely to have a proximal cancer. (21) This is not yet recommended in routine clinical practice. Instead, referral pathways are used to risk assess and stratify patients for either testing (mainly colonoscopy) or a watch and wait approach.

Diagnosis of CRC in symptomatic patients

There are defined CRC diagnosis pathways published by the National Institute of Clinical Excellence (NICE), as well as other organisations, for patients with symptoms that might suggest CRC.

The 2WW pathway allows patients with high-risk symptoms to receive a referral to a specialist or direct access colonoscopy (centre dependent) within two weeks, under the NICE NG12 guidelines. (20) Qualifying symptoms are those aged 40+ with weight loss and abdominal pain, aged 50+ with rectal bleeding, aged 60+ with iron deficiency anaemia (IDA) or change in bowel habit (CIBH), or those with occult blood in their stool. The 2WW pathway should be considered in those with a rectal or abdominal mass, or for those aged <50 with

rectal bleeding plus any of: abdominal pain, CIBH, weight loss or IDA. (20) In addition to this there is provision for referral for patients with lower risk symptoms, if they have a positive faecal immunochemical test (FIT), under the newer (2017) DG30 guideline. (22) This is for symptomatic patients at 'low risk but not no risk' i.e. those at a low probability of having CRC who did not meet the criteria above for urgent 2WW suspected cancer referral. (22) Qualifying symptoms are not age related and include anyone without rectal bleeding but who has other unexplained symptoms. These patients should have a FIT. If the FIT is positive, then the patient can be referred under the 2WW pathway. The FIT requires sampling of a single stool, and uses antibodies to detect the presence of human haemoglobin (Hb). The premise is that CRCs and high risk (precancerous) polyps can cause blood loss from disrupting the integrity of the colorectal mucosal wall. Four FIT systems are fully automated, and provide a quantitative measure of Hb, allowing selection of a threshold of positivity to fit specific circumstances. The sensitivity and specificity of FIT therefore varies according to the chosen threshold for positivity, and according to the patient groups it is being used for.

A 2017 systematic review assessed the effectiveness of very low threshold FIT for primary care setting triage of patients with low risk symptoms (DG30 group of patients). Sensitivities for CRC were reported as 92.1%- 100%, with specificities of 76.6%-85.8%. (23, 24) The FIT has been shown to correctly rule out CRC and avoid colonoscopy in 75% of this low-risk but symptomatic group of patients. (25) FIT was also found to be cost-effective in primary care settings, in those with lower abdominal symptoms who are at low-risk for CRC (24). This shows FIT as an effective triage tool for this low-risk group of patients. Interestingly though, if used in a high-risk-for-CRC symptomatic group (those fulfilling the current 2WW NG12 referral pathway for urgent testing (20)), the sensitivity of FIT is lower at 84.85%, with a specificity of 81.28%. (23) FIT is therefore not adequate to replace colonoscopy for ruling out CRC in the NG12 group, due to its CRC miss rate, yet it is adequate for triaging those with lower risk symptoms (DG30). This is in line with current UK practice, where FIT does not feature in the 2WW referral of patients who fulfil NG12 criteria, but is used for triaging those with lower-risk symptoms, under DG30 guidelines.

Other diagnostic options have also been suggested in symptomatic patients. Multi-target stool DNA testing has been postulated as a tool for use in diagnosis of CRC. DNA tests can include quantitative molecular assays for *KRAS* mutations, aberrant *NDRG4* and *BMP3* methylation, and β-actin, plus a Hb immunoassay. (26) However,

when compared with the FIT, in a large multicentre study, a multi-target stool DNA test detecting the aforementioned mutations showed better sensitivity (92 *vs.* 73%), but lower specificity than the FIT (90 *vs.* 96%) for CRC detection. (26)

Diagnosis of CRC in asymptomatic patients

Another method of diagnosis of CRC is via screening of asymptomatic patients. The rationale is that cancer can be prevented completely by the removal of polyps, as pre-malignant lesions, and that early cancers can be detected before they become large/invasive enough to cause symptoms. The earlier that CRC is detected, the higher chance it has of being curable with surgical resection and/or chemotherapy. (10) The BCSP in England is a nationwide programme, started in 2006 targeting all adults aged 60 to 74, every 2 years. It has been very successful in preventing and detecting cancers. Until recently the BCSP has relied on detection of early-stage cancer using the guaiac faecal occult blood test (gFOBT), which has been demonstrated in trials to reduce CRC mortality (approximately 25% reduction in CRC mortality in those attending for screening). (27) The gFOBT has a variable reported sensitivity for CRC detection of between 12.9%-79.4% and specificity between 86.7%-97.7%. (28-35) The newer FIT is superior to the gFOBT, with respect to sensitivity and positive predictive value for both adenomas and CRCs. (36, 37) The FIT has been extensively tested; and is now used by many CRC screening programmes. (38) In a recent systematic review and meta-analysis, the pooled sensitivity and specificity of FIT to detect CRC in a screening population was 69% and 94% respectively, using thresholds of 6-40µg Hb/g of faeces. (39) Both the gFOBT and FIT rely on the detection of blood in stool as an indication that there could be CRC present. The gFOBT works by a chemical reaction of the guaiac-test paper and hydrogen peroxide developer with the pseudoperoxidase activity of the haem component of Hb in stool samples to produce a blue colour change if blood is detected. However, a positive result can occur due to high meat intake (animal Hb) in the diet or due to certain medications. The FIT, however, is more accurate because it uses antibodies that only detect the globin portion of human Hb, which degrades as it travels down the gastrointestinal (GI) tract. It is therefore less likely to detect upper GI bleeding than the gFOBT. (22) The FIT is therefore now used preferentially (implemented in England in June 2019), both in the UK and other European countries (36), although it is more expensive. (40) Of course both the gFOBT and FIT could test positive if there is blood in stool for another reason such as inflammatory bowel

disease (IBD), haemorrhoids or diverticulitis. The sensitivity and specificity of the FIT is also altered by the quantitative threshold that is chosen to represent a positive result (as introduced above in the symptomatic section). If a low cut-off is used then the FIT will detect a higher number of cancers and colorectal polyps, but may also have a higher false-positive rate, where the capacity of the NHS's endoscopy services may not be able to cope with extra procedures required. A threshold of 120 µg of Hb per gram of faeces is currently the cut-off value for screening in England, but there are plans to reduce this sensitivity threshold. (41)

The flexible sigmoidoscopy screening programme (bowel scope) was also rolled out across England in the BCSP, since 2013, aiming to perform a one-off flexible sigmoidoscopy in all at age 55 years. Results from a 2017 randomised controlled trial (RCT) in 170,432 patients who were assigned to either receive a one-off screening flexible sigmoidoscopy or no intervention showed that CRC incidence was reduced by 26% in the intervention group, and CRC mortality was reduced by 30% after 17 years of follow-up. (13) However, this screening test is being phased out in 2021 due to lack of sufficient resources to deliver the service consistently across England (it was never introduced in Scotland, Wales or Northern Ireland). This decision has been made by NHS England because volume of workload in endoscopy units has increased, particularly with introduction of the FIT into urgent cancer pathways, and there are insufficient trained endoscopists to do the volume of endoscopic procedures required. This is despite NHS England developing a clinical endoscopist training programme to try and address this issue.

Prevention of CRC

The BCSP addresses prevention of CRC as well as early diagnosis of CRC, because colonoscopy leads to detection and removal of colorectal polyps, which may be precursors to CRC. There are however minimal data on the sensitivity of the newer quantitative FIT for polyps. Studies that report on high grade dysplastic lesions often combine the group with CRC patients, making polyp detection rates hard to measure. (23, 24) For low-grade-dysplastic colorectal polyps (less likely to become cancerous), sensitivity of FIT has been reported as 30.6% for all comers to the 2WW. (23) This was however tested in a symptomatic 2WW population, not necessarily representative or appropriate for assessing polyps, as polyps rarely cause symptoms. A recent systematic review and meta-analysis reported the sensitivity of FIT for

advanced adenomatous polyps in an average-risk-of-CRC screening population as 27%, where the FIT cut-off threshold was greater than 20 μg/g faeces (specificity was high at 95%). (42) It makes sense that the sensitivity of FIT (or gFOBT) would be low for polyp detection, as polyps do not always bleed. Additional biomarkers could be revolutionary in the prevention of CRC if they were able to detect polyps, particularly if they could differentiate high-risk polyps in particular.

1.1.2 Barriers to early diagnosis

Work has been done to promote awareness of CRC symptoms in order to improve early diagnosis, notably with the "be clear on cancer" campaign (43), but initiatives to reduce delays to diagnosis have been largely unsuccessful (43, 44). This is partly because symptoms of CRC may be non-specific and mimic a multitude of more common benign bowel pathologies, suggesting why many patients who present with CRC as an emergency have already seen a GP about their symptoms. (45-47)

The BCSP has been shown to decrease mortality and has been a huge success. (48) However, stool tests for detection of blood have their disadvantages as a screening tool for cancer and high-risk polyps, as on the one hand not all cancers and polyps bleed and on the other hand other conditions can also cause bleeding. 33% of the first 17,518 gFOBT-positive patients who attended for colonoscopy as part of the BCSP had diverticular disease as the sole finding, and 15% had haemorrhoids only. (48) Also, there is a false positive rate, where patients are found to have a normal colonoscopy. False positive rates arose in 1% of the population in the Nottingham (49) and Funen (50) RCTs, and up to 9% in the Minnesota RCT (51) (all using a positive gFOBT result to refer to colonoscopy). Much time and resources can therefore be spent doing an invasive investigation on patients who may not have needed the test. This causes undue anxiety and puts patients at a not insignificant risk. The FIT test is more accurate, so there may be fewer false positives, but the number with benign conditions causing bleeding will be unaffected. Another concern with the screening tools used is that the gFOBT or FIT may not be sufficient to pick up proximal cancers; 77% of CRCs found in the first 17,518 colonoscopies (from 2006 to 2008) were distal, with only 14% being proximal. (48) This was surprising given that 28% of CRC in men and 38% of CRC in women diagnosed between 2010-12 had proximal tumours (transverse colon to caecum). (1) It may be that occult blood is not detectable from proximal cancers until the cancer is more advanced. (52)

The gFOBT or FIT may therefore be inadequate as a screening tool for these patients.

Another problem with any stool test is uptake. There was only 54% uptake with the gFOBT used for screening when it was first introduced (down to 35% in the most deprived areas) (53). In a subsequent English pilot study of the BCSP in 2014, once the screening programme was more established, the uptake was reported as higher at 59.3%. (54) However, even this improved gFOBT uptake was much lower than uptake in the breast and cervical screening programmes (71.1% and 71.4% uptake, respectively between 2017-18). (55) One reason for this was deemed to be the unattractive nature of the stool-based test. (53) Other concerns from patients were a perceived hygiene risk, complicated instructions, and the impersonal nature of the process (not GP led). (53) Some of these problems are likely to be addressed as the simpler FIT becomes commonplace; shown by the uptake of the FIT in an English BCSP pilot study in 2014 reported as 66.4%. (54) The unattractive nature of asking patients to collect their stool, however, still remains a barrier. (40, 56) Furthermore, both FIT and gFOBT results may be influenced by drugs such as anticoagulants. A cohort study conducted within the Scottish arm of the national CRC screening pilot on 846 subjects aged 50–69 years showed that taking anticoagulant medication (aspirin, COX-2 inhibitors, other non-steroidal anti-inflammatory drugs or warfarin) at the time of gFOBT was associated with a statistically significant 6.4% increased rate of negative colonoscopy. (57)

Another barrier to early detection is cost. Both screening and symptomatic diagnostic patient pathways are expensive because the required gold-standard diagnostic investigation is an invasive test done in hospital (colonoscopy), by an experienced medical professional. Even if UK incidence rates of CRC remain constant, with an aging population the number of cases and treatment costs are set to double in the next 50 years. This has both cost and work force implications.

1.1.3 Summary: the need for new biomarkers for colorectal cancer

In summary, when used, current pathways appear to work well with regards to detection of CRC in both symptomatic and asymptomatic patients, even if burden on resources is great. The NG12 NICE diagnostic criteria provide a framework for referral (20) in high-risk patients, and the FIT triages lower-risk patients (DG30 criteria (22)) as to whether they need a colonoscopy or not with good sensitivity for CRC. However, the DG30 pathway accounts for only 13% of all of the referrals for suspected CRC. (23) In addition, the FIT also

does not reliably detect early neoplasia or polyps, particularly relevant to the screening cohort for CRC prevention. It also remains the case that most symptomatic patients who have a colonoscopy do not have cancer, reported as up to 42% in one study of 2WW patients. (17, 23) Data from the NICE clinical effectiveness review suggested that 200 of every 1000 patients tested by the FIT do not have CRC but are referred for colonoscopy (false positive results). (58) These findings show that the pathways could be improved with regards to resource allocation and prevention of unnecessary colonoscopies. Additional biomarkers are needed to supersede/enhance FIT results if we are to decrease the number of colonoscopies done in the UK, without missing cancers. (23) For symptomatic patients, a new diagnostic test could streamline NICE guidelines for colonoscopy and may save resources whilst allowing endoscopists to prioritise higher-risk patients. For asymptomatic patients, an ideal screening tool would be one that detects both CRC and high-risk polyps, yet with high enough specificity to reduce unnecessary procedures.

1.2 Biomarkers for colorectal cancer and adenomatous polyps

Carcinoembryonic antigen

The main prognostic biomarker used in routine clinical practice is carcinoembryonic antigen (CEA), a high-molecular-weight glycoprotein, discovered in 1965. CEA is expressed in embryonic tissue and colorectal malignancies. It is measured in patients known to have CRC to indicate cancer progression. It can also indicate recurrence after surgery. It is not used as a diagnostic test, mainly because high CEA levels are not specific to CRC. Other conditions such as IBD, pancreatitis, liver disease and other malignancies and inflammatory conditions can cause it to be raised. (59, 60)

Proteomic approach for biomarkers

There are no blood-based diagnostic biomarkers in routine clinical practice, but there are a number of promising studies which have found proteins that are potentially diagnostic for CRC (see *Table 1-1* for summary). Liquid chromatography mass spectrometry (LC-MS) is the most common technique used, both for CRC tissue and blood (serum and plasma). A systematic review of literature of biomarkers for CRC suggested that panels of serum protein biomarkers can provide sensitivities and specificities above 85% for all stages of CRC, with a

positive predictive value of 72%. (61) Protein assays specifically for autoantibodies have also been investigated as potential biomarkers, where tumour specific autoantibodies may be produced as an autoimmune response to tumour-associated proteins. This is an area of continued research.

Only one of the studies listed below in Table 1-1 investigates proteomics-based biomarkers for adenomatous polyps, where various proteins were highlighted that could potentially differentiate patients with adenomatous polyps from those with CRC or normal colonoscopies (62).

Table 1-1 Recent studies using proteomic-based approaches to find CRC biomarkers

Biomarker	Sample type	Sensitivity	Specificity	Year and reference
Actin beta-like 2	CRC tissue			2017 (63)
Dipeptidase 1	CRC tissue			2017 (64)
Olfamectomedin-4, Kininogen-1, Transport protein Sec-24	CRC tissue			2017 (65)
Cyclophilin A, Annexin A2, Aldodase A	CRC tissue			2016 (66)
Leucine-rich alpha-2-glycoprotein 1, Epidermal growth factor receptor, inter-alpha-trypsin inhibitor heavy-chain family member 4, hemopexin, and superoxide dismutase 3	Blood	70%	89%	2020 (67)
Mannan binding lectin serine protease 1, osteopontin, serum paraoxonase lactonase 3, transferrin receptor protein 1, and amphiregulin	Blood			2019 (68)
Serine/threonine kinase 4	Blood	92.3% (in combination with CEA and gFOBT)	100% (in combination with CEA and gFOBT)	2017 (69)
Macrophage mannose receptor 1, S100 calcium-binding protein A9	Blood			2016 (70)
Alpha-1-antitrypsin, Alpha-1 antichromotrypsin, antithrombin-3	Blood	95%	95%	2016 (62)

*where boxes are empty, no sensitivities or specificities were reported.
Table contents adapted from a table by Alves Martins et al (2019). (64)

Genomic approach for biomarkers

A systematic review in 2014 evaluated DNA and mRNA biomarkers for CRC and colorectal polyps. (61) The overall sensitivity for detection of CRC by faecal DNA markers ranged from 53% to 87%, from 7 papers from 2007-2013. (61) Adenoma detection sensitivity

ranged from 17% to 67%. Combinations of DNA markers tended to increase the sensitivity and specificity of detection of both CRC and adenomas. (71)

Two papers reviewed faecal mRNA markers, namely COX-2 mRNA and MMP7 mRNA (alone or in combination), suggesting sensitivities ranging from 38% to 96% for CRC. Sensitivity for adenomas was reportedly high at 90%, when both markers were used in combination. (71)

Blood/plasma DNA markers (alone and in combination) have also been studied, with sensitivities ranging from 30% to 94% and specificities exceeding 46%, from 8 papers. (61) Four studies also investigated plasma microRNA in circulating tumour cells; combining miR29a, miR92a and miR760 gave a sensitivity of 83% and specificity of 93% for CRC. (72, 73) Lee *et al* (74) reported an adenoma detection sensitivity of 74.7% and specificity of 91.3% for a 4 gene panel—APC, MGMT, RASSF2A, and Wifi1. This would not be an adequate screening tool as "approximately half a million patients from the screening population would have an adenoma missed" with this method. (74)

Overall with regard to the use of biomarkers for screening for CRC, the 2014 review suggested that panels of DNA (74) or miRNAs (72) seemed to be most promising, however validation studies on larger numbers of patients are needed, and that faecal biomarkers may face similar hurdles to the gFOBT with regards to patient uptake. (61)

1.3 **Volatile organic compounds**

1.3.1 Definition and background

Volatile organic compounds (VOCs) are organic chemicals that evaporate at room temperature. VOCs are extensively measured in the food industry, flavour industry and in counter-terrorism. (75-77) Gas chromatography mass spectrometry (GC-MS) is considered "gold standard" for measurement. VOCs emitted from the human body have been of interest for centuries; bad breath and disease state was described by Hippocrates in 420 B.C. Foetor hepaticus as a sign of liver disease and the fruity odour of a patient with diabetic ketoacidosis are still taught in medical schools today. In 1971, Nobel prize-winner Pauling reported that breath and urine contained approximately 250 and 280 VOCs respectively in normal human subjects. (78) It is now known that breath can contain up to 3000 different VOCs, at concentrations ranging from parts per trillion (ppt) to parts per billion (ppb). (79) In the past few decades there has been increasing interest in the study of VOCs for disease states

including cancer, with a more widespread acknowledgment of potential medical applications of VOCs as a diagnostic tool. In 1983 a seminal paper was published naming disease-specific VOCs. (80) More recently, the study of canine scent detection has brought VOC detection in to the headlines. A Labrador dog was trained over 4 years to identify a stool sample from CRC patients from 4 other non-cancer control samples. Samples were watery stools, and the dog was rewarded after correct detection. Over multiple samples, canine scent detection gave a sensitivity of 91% and specificity of 99%. (81) This is better or comparable with any CRC biomarker or screening tool in use today. (61) These findings renewed interest and encouraged enthusiasm into the study of potential clinical applicability of VOC emission from humans.

1.3.2 Mechanism of production

The study of human VOCs is complex, partly because the source of a VOC is often unknown and/or could be from multiple possible processes, and causation is hard to prove, hence a confirmed mechanism from source to the development of disease may be difficult to ascertain. Original sources of VOCs include ingestion (food and medications), physiology (the way the body produces, distributes and processes VOCs) and metabolism of intestinal bacteria (where the presence of disease may alter the microbiome and any subsequent metabolites). (82) In addition to this, concentrations of VOCs may fluctuate rapidly and by a large magnitude in response to changes in human metabolism caused by changes in exertional activity, ingestion of food (timing and type of food), smoking and inflammatory processes. (83) VOCs produced from a cancer site within the body travel within blood to reach pulmonary alveoli where they are exhaled, permitting an objective quantitative measurement. (84, 85) Because of transport via blood, they are also present in urine, faeces, sweat, breast milk, saliva and tissue. (85) Work from the Imperial College London VOC group has established that intestinal tumours are a rich source of VOCs that are transported to the lungs via the blood stream before transferring across the alveolar-capillary membrane to be exhaled via breath. (86) This study compared the concentration of target volatile fatty acids from in-vivo oesophageal tumour headspace (endoluminal air, where the concentrations were found to be highest) to the concentrations in bronchial breath from an endotracheal tube and from exhaled mixed breath. Bronchial breath excluded any oesophageal air that may be present in mixed breath. The equivalence of the concentration between mixed breath and

bronchial breath showed that the origin of the VOCs was indeed the lungs, and by inference, from the systemic circulation. (86) This is an important finding and driver for further work on biomarkers in breath for intestinal disease.

Once detected VOCs must be identified, where the class of VOC reveals its biochemical origin and therefore mechanism of its production, and so the underlying pathophysiological cause may be hypothesised. Alkanes, alcohols and aldehydes may be produced by reduction of hydroperoxide by the cytochrome p450 pathway as part of lipid peroxidation. (82) Cancer and inflammatory pathologies are known to boost oxidative stress, driving lipid peroxidation via the CYP450 enzymes, leading to reactive oxygen species and free radical production from the mitochondria. Circulating reactive oxygen species and free radicals target polyunsaturated fats and proteins where the resultant volatile alkanes and aldehydes can be detected in breath and other media. (85, 87) This alteration in VOC production in patients with cancer postulated to relate to oxygenation of cell membrane-based polyunsaturated fatty acids may result from genetic and/or protein mutations within tumour cells, and lead to the increased relative prevalence of reactive oxygen species. (88, 89) Alcohols may also come from ester hydrolysis by lipases. Aldehydes tend to originate from amino acid and carbohydrate metabolism, and can also be affected by diet. (90) Ketones may be produced from increased fatty oxidation or protein metabolism that occurs during stress, resulting in the formation of ketone bodies. (91) This is demonstrated by the fact that starvation states can be indicated via the breath by the level of acetone present, as a result of increased protein metabolism. (92)

Cancer is widely known to alter metabolism. The Warburg effect is described, whereby high cancer-cell proliferation rates render the tumour cells hypoxic using the conventional oxidative phosphorylation pathway, so cells begin to convert to the use of a glycolysis pathway to keep up with energy demand. (93) Glycolysis leads to lactic acid formation, where the newly acidic environment breaks down the basement membrane of surrounding tissue and protects the tumour from the immune system. (94) Aromatic compounds and nitriles in breath may indicate exogenous sources, such as exposure to tobacco smoke, pollution, alcohol or radiation. They may be carcinogens in themselves however, so should not be disregarded if found to be higher in cancer patients, as they may be both diagnostic and causative. (87) Also, an exogenous source compound may also be endogenously produced. An example of this is ethanol, which is widely present in food and drink, but is also produced

by lung tumour tissues and exhaled by lung cancer patients and is also produced by intestinal bacteria such as *Escherichia coli*. (83)

1.3.3 Volatile organic compounds in current clinical practice

VOCs are transported in blood, urine, faeces, breast-milk, breath, and are present in all tissues of the human body. (95) However, the analysis of exhaled breath for detection and monitoring of disease is particularly attractive, as it is not only non-invasive but is extremely easy to perform, and likely to be highly acceptable to patients. (96) Routine clinical applications of breath testing include the roadside alcohol breathalyser, ^{13}C urea breath test for Helicobacter pylori, nitric oxide in asthma, and the hydrogen/methane test for intestinal bacterial overgrowth (97-101).

1.3.4 Breath testing for cancer using volatile organic compounds

In 2018, the Imperial College London VOC group performed a systematic review of literature and a meta-analysis looking at the diagnostic accuracy of VOCs for cancer diagnosis in breath, excluding sensor technologies. (102) The search yielded 63 publications, looking at lung, breast, gastroesophageal, colorectal, oral cavity, head and neck, mesothelioma, thyroid, liver, ovarian, prostate and laryngeal cancers. Most studies used GC-MS to analyse the breath samples, and most compared the cancer group against healthy controls. A total of 253 VOCs were reported in association with cancer diagnosis in different tumour sites; mainly "hydrocarbons (aromatic and aliphatic) and oxygenated compounds (namely, aldehydes, alcohols, phenols, carboxylic acids, ethers, and furans) and less frequently nitrogen-, sulphur- and halogen-containing compounds". (102) Specific VOCs that were linked with cancer (in order of frequency) were 2-butanone (14 studies), 1-propanol, nonanal, isoprene, ethylbenzene, 4-methyl octane, 3-hydroxy-2-butanone (each in 8 studies), acetone, toluene, ethanol, pentanal, heptanal and pentane (each in 7 studies). The overall pooled analysis of outcomes found a sensitivity of 79% and specificity of 89% for breath testing for cancer in general. (102) The Imperial College London VOC group has also been developing a breath test for the detection of oesophagogastric adenocarcinoma, from biomarker discovery to clinical trials. The assessment of VOCs within exhaled breath by selected ion flow tube mass spectrometry (SIFT-MS) (103) was followed by the construction and validation of a diagnostic

model using a panel of 9 VOCs in 225 patients with an area under the Receiver Operating Characteristic (ROC) curve of 0.92 and 0.87 for the model and validation subsets respectively (104). Cross-platform chemical validation was performed using GC-MS. A National Institute for Health Research (NIHR)-funded multicentre blind validation study of VOC breath analysis in 396 oesophagogastric patients (105) showed a sensitivity of 86% and specificity of 80%. In parallel with this work, the group has investigated the exhalation kinetics and molecular drivers of VOCs in oesophagogastric adenocarcinoma. (106-108) Chapter 2 reviews the literature in detail regarding the use of VOCs specifically for colorectal diseases including CRC.

1.4 **Rationale and aim of PhD project**

Rationale for PhD project

There is a need for a more sensitive and specific test to aid in the goal of early detection of CRC, as demonstrated by the fact that still only 10% of patients are detected in the asymptomatic phase, when CRC is most likely to be curable. NHS resources are scarce and the burden on endoscopy units great, hence a more streamlined approach to triaging those most likely to have cancer would be of great benefit. An intermediate test could not only reduce the number of non-CRC patients qualifying for the current urgent 2WW pathways, but would move the appropriate patients who are currently on routine referral pathways to urgent lists. FIT is one answer to this problem, but as discussed, it has a false positive rate, and most importantly a low but significant miss-rate, as well as an imperfect uptake rate. (22, 23). Breath testing has ideal characteristics for a triage test being non-invasive, simple to perform, and likely to be universally acceptable to all. (96) These characteristics may lead to good patient uptake. A primary care setting would likely be the most useful and appropriate setting for such a test. The Imperial College London VOC group, among others, has already produced promising work on the use of breath testing to detect CRC, providing motivation for this study. (109)

Possible roles of a future breath test could include:

1) An adjunct to the NG12 2WW criteria, aiming to reduce the referral burden by ruling out CRC.
2) As a replacement or adjunct to the FIT in those who have lower risk symptoms (DG30 fulfilling patients). This could streamline NICE guidelines for colonoscopy and may

save resources and allow endoscopists to prioritise those who are in fact higher-risk for CRC.

3) As a screening tool, if results were superior to that of the FIT (taking into account acceptability and potential cost also).

It is worth mentioning though that a breath test would never replace colonoscopy, as many patients with symptoms need a colonoscopy for diagnosis of other pathology; 28.9% of people with a 'false-positive' FIT for CRC were found to have other treatable and significant bowel pathology, such as IBD or high-risk adenomata. (58) A breath test would purely streamline the current pathways.

Aims of PhD

My PhD aims to forward the work already done by the Imperial College London VOC group and address some of the challenges of breath collection by:

1) Reviewing the existing literature on the use of VOCs for the diagnosis of colorectal disease including CRC.
2) Developing and outlining a system to ensure quality control of breath samples, for the entire breath testing process.
3) Performing a large multi-centre study to investigate the use of VOCs present in exhaled breath for the prediction of CRC and adenomatous polyps.
4) Performing a large multi-centre study investigating the feasibility and acceptability of breath testing in primary care.

I will elaborate on these broad aims in the following four dedicated chapters.

2 CHAPTER 2: SYSTEMATIC LITERATURE REVIEW OF VOLATILE ORGANIC COMPOUNDS IN COLORECTAL DISEASE

I completed a systematic review of the literature on VOC testing for colorectal disease, including both CRC and non-cancerous colorectal disease, using breath, blood, faeces and other biological matrices.

2.1 **Methods**

An electronic search (title and abstract) was performed using the Embase, MEDLINE, Global Health, Health Management Information Consortium, Books@OVID, Journals@Ovid and Imperial Journals@Ovid databases through the OVID platform, (1946 to 20th May 2020), in accordance with the recommendations of the Cochrane Library.

2.1.1 Search strategy and selection criteria

The search terms *cancer*, *neoplasm*, *malignancy*, *disease, disorder, pathology, volatile organic compound, VOC, colo*, rectum, intestin*, gastro** and *GI* were used in combination with the Boolean operators AND and OR. The search was limited to human studies published in the English language. Conference abstracts were excluded because of the high probability of incomplete data, precluding a thorough review. Full details of the search strategy are provided below in Table 2-1.

Table 2-1 Search strategy used in Ovid online search engine

	Searches	**Number of results**
1.	Cancer OR neoplasm OR malignancy	10240063
2.	Volatile organic compound OR VOC	41653
3.	Colo* OR rectum OR intestin*	6901751
4.	Disease OR disorder OR pathology	25348582
5.	Gastro* OR GI	4946127
6.	1 OR 4	30459075
7.	3 OR 5	10129678
8.	2 AND 6 AND 7	2705
9.	Remove duplicates from 8	2135
10.	Limit 9 to English language	**2126**

Full texts of potentially relevant articles were retrieved to determine eligibility for inclusion in the review. Studies were included if they met the following criteria: (1) trials that analysed endogenous VOCs for the diagnosis or assessment of histologically confirmed chronic colorectal diseases; and (2) studies that included VOC identification. Studies were excluded if: (1) an exogenous substrate (intravenous, oral, or inhaled) was administered before sampling; (2) sensor-based pattern recognition technology was used in the absence of VOC identification; (3) the colorectal disease in question was not a chronic pathology with histological diagnosis, such as transient infectious diseases, irritable bowel syndrome (IBS), post-operative ileus; or (4) studies were animal studies. The reference lists of included studies were also reviewed to identify any additional relevant studies.

For each study, data related to disease targeted, nature of biofluid/sample medium used, discriminative VOCs, analytical platform, and sensitivity, specificity, and the area under the ROC curves derived from diagnostic models was extracted.

2.1.2 Quality Control

Each diagnostic study identified was evaluated for quality of reporting using the Standards for Reporting of Diagnostic Accuracy Studies (STARD) checklist. (110) This is an evidence-based checklist of 30 items addressing bias, variability, and limitations of the applicability of results to other studies. A high STARD score represents a diagnostic accuracy study with high standards for completeness and transparency of reporting.

A modified version of the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) tool was also used to assess the quality of the included studies. (111) QUADAS-2 consists of 4 domains, including patient selection, index test, reference standard, and flow of patients through the study. This modified version has been used by other reviews looking at phase 1 biomarker discovery studies using VOCs. (102) Full details of the modified QUADAS-2 tool used are presented in Appendix 1. The modified version places particular importance on the inclusion of positive controls, internal and/or external validation of results, assessment before therapeutic intervention, and suitable reproducibility and sensitivity of the index test. The modified tool has previously been independently piloted. (102)

2.2 **Results**

The literature search yielded 23 results that met the inclusion criteria, where 4 had been found by review of the reference lists of other studies. See the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram for detail of the literature search, in Figure 2-1. All reports constituted phase 1 biomarker discovery studies. Of these, 10 related to benign diseases of the colon (IBD in all cases), 12 related to CRC, and 1 related to adenomatous polyps (although the title and abstract may lead the reader to believe that the study is about CRC, rather than *risk of* CRC). (112) The number of patients with the reference disease (IBD, polyps or CRC) included in those studies ranged from 16 to 217, where the maximum number used in the statistical analysis was 187.

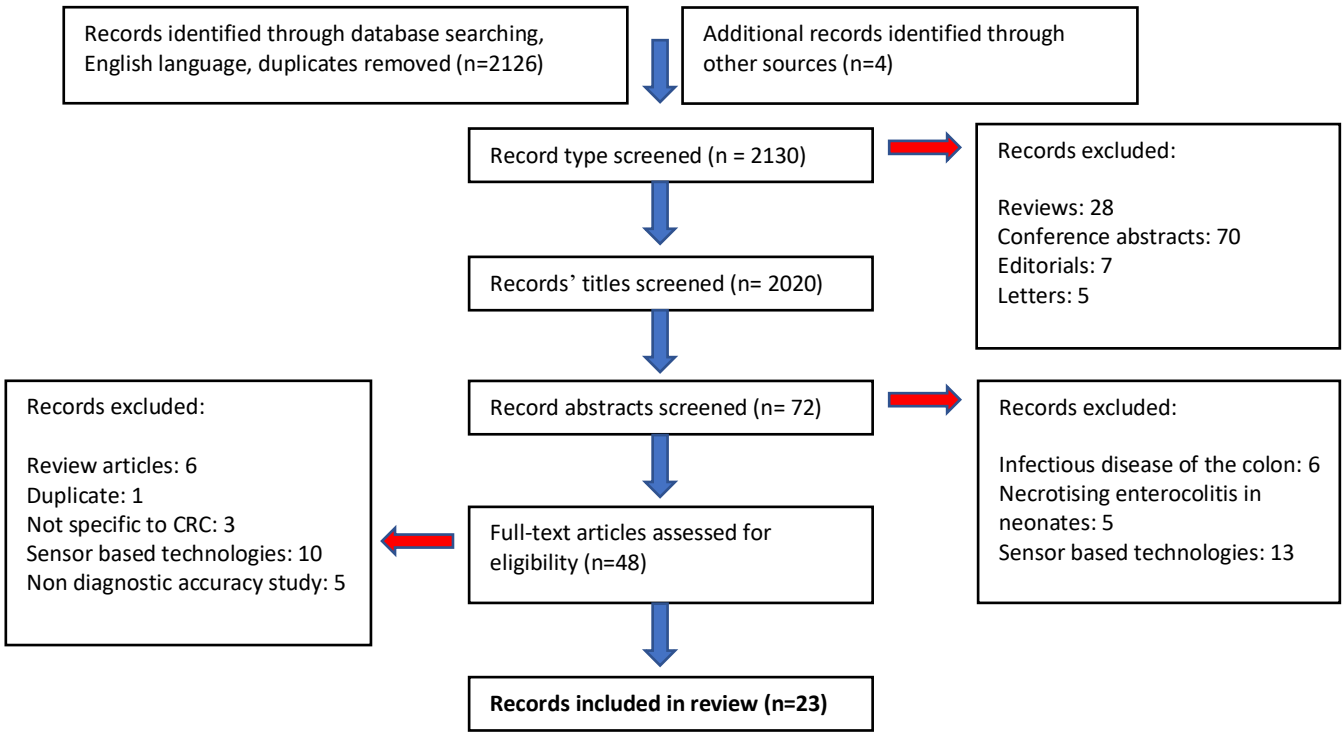


Figure 2-1 PRISMA flow diagram of literature search

2.2.1 Benign diseases of the colon

Of the 23 studies included in this review, 10 related to benign diseases of the colon and were reviewed in full; VOCs were analysed from breath (n=7, where 3 were paediatric studies) and faeces (n= 3). All studies used patients with IBD as the disease cohort, either with Crohn's disease (CD) in isolation (n=1), ulcerative colitis (UC) in isolation (n=1) or IBD split into CD and UC groups (n=8). One study's objective was to discriminate IBS patients, but included UC and CD patients as comparators, so was included in the review. (113). In all studies the comparators (controls) were asymptomatic healthy patients (n=9), patients with gastrointestinal symptoms but no structural or histological pathology including IBS (n=3), or those in remission with inactive IBD (n=2). Some studies used more than one of these comparator groups (n=4). See the IBD subsection of the summary of studies below in Table 2-2 for full details.

All studies showed a significantly different VOC profile between the IBD group(s) and non-disease controls. A total of 99 different key discriminatory VOCs were identified across the studies. Discriminating VOCs for IBD that were found to be significant by more than one study were acetone (n=2), pentane (n=2), ammonia (n=2), acetic acid (n=3), 3-methyl-1-butanol (n=2), alpha-pinene (n=3), heptanal (n=2), 2-piperidinone (n=2), 6-methyl-2-heptanone (n=2), butanoic acid (n=2) and 1-propanol (n=2). Two of the acetic acid studies, one of alpha-pinene studies and one of the 3-methyl-1-butanol studies were from paediatric cohorts. The sensitivity of breath testing for IBD (CD and/or UC) vs controls (healthy, IBD in remission or IBS controls) ranged from 81% to 96%. Specificity ranged from 69% to 94%. See summary of studies for details of studies, VOCs and diagnostic accuracy results, below in Table 2-2.

The STARD score for reported IBD studies ranged from 8 to 23, with a mean (standard deviation) of 16.1 (4.4). The details of STARD for each study are also provided in Table 2-2. Assessment of biases and applicability of outcomes using QUADAS-2 are detailed in Appendix 1 and 2. Major areas of bias were (1) lack of a positive control group, where 'healthy controls' were used instead; (2) the lack of a validation set to confirm diagnostic accuracy for the index test; (3) failing to report the reference standard test (particularly for the control group, where there was often no endoscopic or histological confirmation of their 'healthy control' status and; (4) failing to acknowledge the interval between index and reference tests, hence in some studies it was unclear whether they were retrospective or prospective.

2.2.2 High-risk colorectal polyps

One study compared the VOCs from faeces in a high-risk for CRC group (n=31) compared to a low-risk for CRC group (n=31). (112) On review of the demographics it became clear that the high-risk group were all high-risk adenomatous polyp patients, with no CRC patients included. Although this was a small study with multiple areas of potential biases (see detail of studies including the STARD checklist, Table 2-2), it was interesting that 4 of the 5 VOCs that were found to discriminate high-risk polyps from healthy controls were VOCs that had been reported for detection of IBD by other studies in this review. (114, 115) These overlapping VOCs were hydrogen sulphide, dimethyl sulphide, ammonia and acetaldehyde. There were no overlaps with compounds found for CRC. No other studies investigating VOCs for detection of adenomatous polyps were found by this systematic review of literature.

2.2.3 Colorectal cancer

Of the 12 included studies on CRC reviewed in full, VOCs were analysed from breath (n=5), faeces (n=1), urine (n=1) and plasma (n=5). All studies used patients with CRC as the disease cohort, however on many occasions the histological classification was unspecified. All but 1 study used 'healthy controls' as a comparator; definitions of this population included patients with a normal colonoscopy (n=5), a non-cancer colonoscopy (n=2), a non-cancer and non-polyp colonoscopy (n=1), lack of GI disorders on clinical review (n=1), and undefined (n=2). One study used the same cohort of CRC patients, after operative removal of their cancer, as paired controls (116). See summary of studies below in Table 2-2.

All studies showed a significantly different VOC profile between the CRC group and the comparator group. Each study found between 1 to 44 VOCs that were predictive of CRC. All are detailed in Table 2-2 below. 13 of these VOCs were independently identified by more than one study (never by more than two studies) as significant for CRC diagnosis: nonanal, decanal, 2-methylbutane, 4-methyl-2-pentanone, 1,2 pentadiene, 2-methylpentane, cyclohexane, methylcyclohexane, 1,3-dimethylbenzene, 1,4-dimethylbenzene, 4-methyloctane, dodecane and 6-t-butyl-2,2,9,9-tetramethyl-3,5-decadien-7-yne. These VOCs were within the aldehyde, alkane and alkene chemical groups, as well as one ketone. 5 VOCs overlapped with those found to be discriminatory for IBD. These were acetone, 1-octene, octane, methylcyclopentane and undecane, again ketones, alkenes and alkanes. Three VOCs

overlapped with those found for other cancer types, see previous section: nonanal, acetone and ethanol.

The reported sensitivity of VOC testing for CRC vs controls was 83% to 100%. Specificity ranged from 76% to 100%. See summary of studies for details of VOCs and diagnostic accuracy results, in Table 2-2.

The STARD score for reported CRC studies ranged from 13 to 24, with a mean (SD) of 17 (3.3), again detailed in Table 2-2-2. Assessment of biases and applicability of outcomes using QUADAS-2 are detailed in appendix 2.

Major areas of bias were: (1) lack of a defined positive control group; where "negative colonoscopy" was used as a control group, there was rarely any detail about any benign or other pathology found, so it was unclear as to whether the control group was a positive control group or not. (2) Lack of a defined negative control group; where 'healthy controls' were used as a comparator, detail was rarely given as to the definition of "healthy", or any detail about any comorbidities. (3) Lack of a validation set in many studies, which would aid in confirming the diagnostic accuracy of the index test. (4) Failure to acknowledge the interval between index and reference tests, hence in some studies it was unclear whether they were retrospective or prospective. (5) Unequal testing of disease and control groups mainly with regards to timing and location of testing; often cancer groups were tested before an operation, whereas non-disease groups were tested in clinics/endoscopy, or it was unclear when/where they were tested. (6) Variability in statistical methods to produce a sensitivity and specificity value; in one study, data from a sensor was used for this calculation rather than the data from GC-MS (also measured) (117), so it was unclear whether the values pertain to the VOCs stated, as sensors are likely to detect different compounds from the GC-MS.

Table 2-2 Detail of all included studies for VOC testing for colorectal disease using breath, blood, faeces and urine.

Author and ref	Year	Study topic	Disease	Controls	Total	Disease number	Analytical platform	Discriminating VOCs	Sens	Spec	AU ROC	Comparison with best result	STARD score
IBD studies													
Rieder *et al (118)*	2016	IBD breath	CD, UC, IBDu, IPAA	Healthy	124	24 CD,11 UC,6 IBDu, 53 HC, 30 IPAA	SIFT-MS	4 VOCs: acetone, acrylonitrile, carbon disulphide, triethylamine	?	?	0.81	IBD v non IBD	13
Hicks *et al (114)*	2015	IBD breath	CD, UC	Healthy	56	20 UC, 18 CD	SIFT-MS	6 VOCs: dimethyl sulphide, hydrogen sulphide, hydrogen cyanide, ammonia, butanal, and nonanal	0.94	0.94	0.86	CD v healthy controls	15
Bodelier *et al (119)*	2015	IBD breath	CD	Remission, Healthy	835	140 active, 135 remission, 110 HC	GC-MS	6 VOCs: isoprene, and acetone were decreased in the active CD group, whereas those of 2,2,4-trimethylpentane, branched saturated C12 compound (RT 1⁄4 16.6 min), 1-butoxy-2-propanol, and heptadecane are elevated.	0.81	0.80	0.80	Active CD v healthy controls	22
Patel *et al (120)*	2014	Paediatric IBD breath	CD, UC	Healthy	117	62	SIFT-MS	3 VOCs: 1-octene, 1-decene and (E)-2-nonene	?	?	0.96	IBD v healthy controls	17
Dryahina *et al (121)*	2017	IBD breath	CD, UC	Healthy	201	136 CD, 51 UC	SIFT-MS and GC-MS	5 VOCs: pentane, hydrogen disulphide, acetic acid, propanoic acid and butanoic acid	?	?	?	IBD vs healthy controls	8

Monasta *et al (115)*	2017	Paediatric IBD breath	CD, UC,	Benign and Healthy	234	34 CD, 33 UC, 65 non-IBD positive controls, 102 healthy	IMR-MS	18 VOCs: methane, acetic acid, ammonia, propene, nitrous acid, acetaldehyde, methylethylketone, other masses unknown ID	0.96	0.69	0.93	IBD v all controls	16
Smolinska *et al (122)*	2017	Paediatric IBD breath	UC	Remission, non-UC colitis	94	72	GC-MS	11 VOCs: cumene, 2,4-dimethylpentane, methylcyclopentane, C14H30 branched, C15H30(pentadecene), 3-methyl-1-butanol, octane, acetic acid, alpha-pinene, m-cymene	0.92	0.77	0.94	UC v remission	23
Walton *et al (123)*	2013	IBD faeces	CD, UC,	IBS, healthy	87	CD 22, UC 20, IBS 26, HC 19	GC-MS	7 VOCs: Butanoic acid, 1-propanol, ethyl ester propanoic acid , ethyl ester butanoic acid , indole, 1-butanol, 3-methyl butanoic acid, phenol	?	?	?	CD v healthy controls (significant compounds only, no model)	11
Ahmed *et al (124)*	2016	IBD faeces	CD, UC	Healthy	326	CD 117, UC 100	SPME GC-MS	8 VOCs: Heptanal, 1-octen-3-ol, 2-piperidinone and 6-methyl-2-heptanone and decane were up-regulated in the active CD group, while methanethiol, 3-methyl-phenol, alpha pinene were down regulated	?	?	?	CD (active) v healthy controls	17
Ahmed *et al (113)*	2013	IBD faeces	CD,UC	IBS, Healthy	249	30 CD, 62 UC, 48 IBS	SPME GC-MS	44 VOCs: Esters of short chain fatty acids, cyclohexanecarboxylic acid and its derivatives were significantly associated with IBS-D groups, aldehydes and ketones were associated with CD	0.94	0.82	0.97	CD v IBS	19

								groups while the three VOMs more abundant in UC were (1-propanol, 2-methyl), (undecane) and (methoxy-phenyl-oxime).					
CRC studies													
Markar *et al (109)*	2019	CRC Breath	CRC	HC with benign colorectal disease, HC with normal colonoscopy	150	50	SIFT-MS	1 VOC: Propanal	0.96	0.76	0.90	CRC v negative control	24
Amal *et al (117)*	2016	CRC, polyps, Breath	CRC, polyps	Healthy controls (negative colonoscopy)	418	CRC 65, polyps 22	GC-MS	4 VOCs: acetone and ethyl acetate (higher in CRC), ethanol and 4-methyl octane (lower in CRC)	0.85	0.94	?	CRC v control	17
Altomare *et al (116)*	2015	CRC Breath	CRC	Paired test, Disease-free (post-op)	48	32	GC-MS	31 VOCs: 1,2 pentadiene, 2-methylbutane, 2-methylpentane, 1-(1-methylethenyl1)-2-(1-methylethyl1)benzenebmethylcyclopentane, cyclohexane, heptane, methylcyclohexane, 4-methyl-2-pentanone, 1-methylnapthalene, 1-octene, octane, 1,2,3-trimethylbenzene, 1,3-dimethylbenzene, 1,4-dimethylbenzene, propylbenzene, beta-pinene, 1-methyl,3-(1-methylethylbenzene), X, 5-	1.0	0.96	0.99	48 CRC v 32 paired post-op	18

								butylnonane, undecane, nonanal, dodecane, decanal, 1-ethyl-1,2,4-trimethylbenzene, 1-ethyl-,2,4,5- trimethylbenzene, 2,3-dihydro-1,6-dimethyl-1H-indene, 2,3-dihydro-4,7-dimethyl-1H-indene, 1,3 dimethyl1-5-(1-methylethyl)benzene, 2 methylnapthalene						
Altomare *et al (84)*	2013	CRC Breath	CRC	Controls (negative colonoscopy)	78	37	GC-MS	15 VOCs: Nonanal, 4-Methyl-2-pentanone, Decanal, 2-Methylbutane, 1, 2-Pentadiene, 2-Methylpentane, 3-Methylpentane, Methylcyclopentane, Cyclohexane, Methylcyclohexane, 1,3-Dimethylbenzene, 4-Methyloctane, 1,4-Dimethylbenzene, A (4-methylundecane, RT = 11·3), B (trimethyldecane, RT = 13·2)	0.86	0.83	0.85	CRC negative control	v	19
Bond *et al (125)*	2019	CRC Faeces	CRC, polyps	Controls (negative colonoscopy)	137	21, 56	GC-MS	2 VOCs: propan-2-ol and ethyl 3-methyl- butanoate	0.88	0.85	0.82	CRC negative control	v	22
Kim *et al (126)*	2019	CRC, Plasma	CRC	controls (?definition)	60	30	GC/GC-MS	5 VOCs: 2,4,6-Trimethyl-1-nonene, 2,4-dimethylhept-1-ene (increased), 2,3,4-trimethyl-hexane, 3-ethyl-	?	?	?	CRC control	v	13

								hexane, 3,5-dimethyl-octane							
Wang *et al (127)*	2014	CRC, Plasma	CRC	Healthy controls (no cancer on endoscopy)	36	16	SPME GC-MS	4 VOCs: found at significantly lower levels in the group of CRC patients than in the normal control group (P < 0.01): phenyl methylcarbamate, ethylhexanol, and 6-t-butyl-2,2,9,9-tetramethyl-3,5-decadien-7-yne. In addition, significantly higher levels of 1,1,4,4-tetramethyl-2,5-dimethylene-cyclohexane were found in the group of CRC patients than in the normal control group (P < 0.05).	?	?	?	CRC negative control	v	14
Wang *et al (128)*	2014	CRC, Breath	CRC	Healthy controls (no cancer on endoscopy)	40	20	SPME GC-MS	9 VOCs: significantly higher levels of cyclohexanone, 2,2-dimethyldecane, dodecane, 4-ethyl-1-octyn-3-ol, ethylaniline, cyclooctylmethanol, trans-2-dodecen-1-ol, and 3-hydroxy- 2,4,4-trimethylpentyl 2-methylpropanoate but significantly lower levels of 6-t-butyl-2,2,9,9-tetramethyl-3,5-decadien-7- yne (P < 0.05).	?	?	?	CRC negative control	v	14
Crotti *et al (129)*	2016	CRC, Plasma	CRC	Healthy controls. UC, HG	117	48	GC-MS	1 main VOC: Decanoic acid could differentiate between CRC and	0.88	0.80	0.82	CRC healthy controls	v	16

				adenoma, breast cancer				healthy/UC/breast cancer controls, but no difference between CRC and high risk polyps. CRC and polyps indistinguishable.							
Tan *et al (130)*	2013	CRC, Plasma	CRC	Healthy controls ("no GI disorders")	203	101	GC-MS and LC-MS	10 metabolites: tricarboxylic acid (TCA) cycle, urea cycle, glutamine, fatty acids, and gut flora metabolism.	0.84	0.92	?	CRC healthy controls	v	13	
Nishiumi *et al (131)*	2012	CRC, Plasma	CRC	Healthy controls (no endoscopic, plasma, imaging or medical interview abnormalities)	122	59	GC-MS	4 VOCs: 2-hydroxybutyrate, aspartic acid, kynurenine, and cystamine,	0.91	0.85	0.90	CRC healthy controls	v	18	
Cheng *et al (132)*	2012	CRC, Urine	CRC	Healthy controls (no endoscopic, plasma, imaging or medical interview abnormalities)	204	101	GC-MS and LC-MS	7 VOCs: citrate, hippurate, *p*-cresol, 2-aminobutyrate, myristate, putrescine, and kynurenate	0.98	0.10	0.99	CRC healthy controls	v	16	
Colorectal adenoma studies															

Batty *et al (112)*	2015	High risk polyps, Faeces	"High risk"	Controls (non-cancer colonoscopy)	62	31	SIFT-MS	5 VOCs: hydrogen sulphide, dimethyl sulphide and dimethyl disulphide, ammonia and acetaldehyde	0.72	0.78	?	High risk polyps v non-cancer	12

Abbreviations used in table: IBD = inflammatory bowel disease, CD = Crohn's disease, UC = ulcerative colitis, IBDu = inflammatory bowel disease unclassified type, IBS = irritable bowel syndrome, HC = healthy control, IPAA = ileal pouch-anal anastomosis, GC-MS = gas chromatography mass spectrometry, SIFT-MS= Selected Ion flow tube mass spectrometry, SPME = solid-phase microextraction, LC-MS = liquid chromatography mass spectrometry.

2.3 **Discussion**

In general, the current literature surrounding VOCs for testing for colorectal diseases largely consists of pilot studies, with relatively low numbers of participants. All are single-centre studies and there are no large RCTs. Methods for VOC collection and analysis remain unstandardised, as is the quality control (QC) element of sampling. Individual laboratories therefore vary as to the complexity and rigor of their QC processes, and may apply QC protocols at different stages in the workflow from equipment preparation to sampling to analysis. The most common analytical platform is GC-MS, where all but one of the CRC studies used this. SIFT-MS was the next most commonly used platform. Urine, breath, faeces and plasma were all used, with different collection methods. Given these differences it is unsurprising that studies came up with different differentiating VOCs, particularly when you also consider that GC-MS studies are mainly untargeted analyses as opposed to SIFT-MS analysis which detects target ions. These variations are also combined with the inherent intra- and inter-individual variability that will always remain a challenge in VOC research. (102) Regarding the variability in the included studies, often gender and age were reported, with detail about endoscopic findings, but rarely was there detail of diet, smoking, exercise, body mass index (BMI), fasting time, co-morbidities or concomitant medications. These have all been documented to affect the global human VOC profile (133, 134), but it is not known if these affect the cancer-specific biomarkers.

Despite all the variation between studies however, they were unanimous in finding that CRC, IBD, and, to a lesser extent, colorectal polyp groups could be clearly differentiated from comparator groups, often with very high sensitivities and specificities. This is very promising for future early stage diagnosis strategies. Largely the compounds found to discriminate CRC were not the same as those found for IBD, with the exception of 5 VOCs (acetone, 1-octene, octane, methylcyclopentane and undecane). The most frequently identified compounds in the literature found to discriminate colorectal pathologies were in the aldehyde, alkane, alkene and ketone groups. Most studies postulated a possible mechanism for the biomarkers that they had found. Common themes included the altered metabolism in inflammation and cancer, as well as alterations in the microbiome. However, as stated in previous reviews, there are few studies designed to investigate the molecular mechanisms and/or exhalation kinetics

(in the case of breath) that lead to the production of VOCs in cancer and other colorectal pathology (102).

Limitations

This systematic review did not include studies that used sensor or pattern recognition technologies, although there are some large studies for CRC using these with encouraging results. (96, 135) This is because these studies cannot report the identity of the volatile biomarkers. The majority of the studies found did not perform external validation of their findings, therefore the reliability and reproducibility of some of the reported results is questionable. Mechanistic studies were not specifically searched for in this review, hence the question of why VOCs may be different between pathologies was not addressed.

Overall, given the small study sizes, variable methods and matrices, and poor quality of some of the studies, there were only three main relevant studies for this PhD of breath testing in CRC. Altomare *et al* used GC-MS to compare VOCs in exhaled breath from 37 CRC patients and 41 control patients. When the resulting model was applied to prospective 25 breath samples (15 with CRC), the overall diagnostic accuracy was 76%. (84)

Amal *et al* used GC-MS analysis to reveal four VOCs that significantly differed in the exhaled breath from CRC patients and the control group (acetone, ethyl acetate, ethanol and 4-methyl octane). This study using 209 patients (65 with CRC) showed that VOC detection had 85% sensitivity and 94% specificity to detect CRC. (117) Patients had been referred for surgery or endoscopy, so it is assumed that they were symptomatic (although this is not explicit). Despite being a small study this is a comparable sensitivity and better specificity as reported for the FIT test in symptomatic patients (a recent systemic review and meta-analysis by Stonestreet *et al* demonstrated a pooled sensitivity of 90% and specificity of 87% for CRC detection by FIT). (39)

A prospective pilot study was performed by our group at Imperial College London to identify discriminative VOCs for CRC (n=150). This was followed by a diagnostic validation study (n=110); a clinical prognostic study following surgery and after tumour recurrence (n=60); cross platform chemical validation using GC-MS (n=30). The diagnostic model had a sensitivity of 85% and specificity of 84% for cancer diagnosis. Following surgery, the biomarker normalised to control level, but levels significantly increased with recurrence. (109)

2.4 **Conclusion**

This review highlights the need for larger studies, rigorous QC procedures and crucially standardisation in the area of VOC detection for colorectal pathology diagnosis (102). Following the discovery of reproducible and robust biomarkers, mechanistic studies are then required to provide plausible explanations.

3 CHAPTER 3: METHODS AND MEASUREMENT OF BREATH COLLECTION, AND ENSURING ROBUST QUALITY CONTROL

3.1 Breath collection systems

Breath is a complex matrix of thousands of VOCs and water, reflecting directly the products of the exchange between the blood stream and the alveoli, via the alveolar membranes. It is also influenced by contact with pathology or bacteria present in the airway or naso-oral cavity. As such it must be carefully captured without risk of escape or contamination, and analysed by an instrument capable of discovering its complex make-up. Mass spectrometry is widely used as it has the ability to separate and identify the components of a complex mix of compounds, is well understood as a process, and is readily available. A variety of breath collection systems exist for VOC analysis. Options discussed here are in the context of clinical studies, given that this the focus of this current PhD, rather than lab-based smaller scale studies.

3.1.1 Online breath sampling

Mass spectrometers such as the SIFT-MS and Proton Transfer Reaction-Time of Flight-Mass Spectrometer (PTR-ToF-MS) permit direct, on-line sampling and analysis. This is because the patient can exhale directly into the inlet of the instrument via a mouthpiece, so that the mass spectra appear on screen in real time. This is described elsewhere. (136) See *Figure 3-1*.

Advantages: Fast and easy to perform for the patient. No contamination, storage or degradation of samples, as there are no intermediate sampling steps or time delays between the exhalation and collection by the instrument. No loss of trace compounds.

Disadvantages: Not feasible for large multi-centre studies, as these delicate instruments cannot easily be moved and clinical patients often cannot plausibly come to a laboratory for sampling. No storage options for breath samples. Direct sampling permits analysis of one exhaled breath only at any one time, whereas other techniques may assess multiple exhalations and thereby detect a higher quantity and concentration of VOCs.



Figure 3-1 Online breath sampling into the SIFT-MS

3.1.2 Breath bags

Bags can be used as a breath reservoir, where collection is straightforward as the patient simply blows through an inlet to inflate the bag, which is then sealed. Multiple types of breath collection bags have been used in the literature, with 3 used by the Imperial VOC group. The nalophan bag is made of double thickness (2 x 25micrometre) Nalophan® (Kalle UK Ltd., Witham, UK) material, where the inlet is a 1mL Luer lok syringe (Terumo, Europe, Leuven, Belgium). The syringe is made up of a syringe barrel which is secured to the inside of the Nalophan® bag with a tight seal, and a syringe plunger which can be re-inserted into the syringe barrel after sampling in order to create an air-tight plug to prevent breath from escaping. See Figure 3-2. Steel bags can also be used, via the same mechanism; patients exhale into a 500-mL steel breath bag (GastroCHECK) via a 1-mL Luer lok syringe (Terumo Europe, Leuven, Belgium). Tedlar® bags (Adtech Polymer Engineering Ltd, Glos., UK) can also be used (a type of fluoroplastic), where capacity size can be 1 to 300L, with a thickness of 50microns. The difference with these is the purpose-made polypropylene hose, valve and septum for the patient to breath in to, instead of a syringe.

Advantages of bags: Breath collection is easy, quick and can be done at the patient's bedside, and is therefore highly acceptable to patients. Minimal number of components make contamination less likely. Storage of breath is possible; up to 6 hours for trace compounds and up to 24 hours for abundant compounds (such as acetone) was reported using Nalophan®

bags, and 24 hours was reported for Tedlar® bags.(137) Longer than this may lead to degradation of the VOCs, and therefore unreliable detection.

Disadvantages: Bags need to be analysed quickly, which may make large studies unfeasible. Bags cannot easily be transported (fragile and bulky), hence precluding multi-site clinical studies.



Figure 3-2 Breath sampling into a Nalophan® bag

3.1.3 Thermal desorption tubes

An alternative option for breath collection is thermal desorption (TD) tubes. These are currently the only commercially available option if longer storage times and transport across different study sites are required, typically in larger or multi-centre clinical studies. TD tubes are stainless steel cylindrical tubes, 89mm long and 64mm in diameter (Markes International, Llantrisant, UK), packed with a Carbograph/Tenax-5TC sorbent, designed to absorb VOCs within breath. The stability of the VOCs in sorbent tubes has been previously evaluated at different temperatures for storage, showing an acceptable stability for approximately one month at room temperature, and 1.5 months at -80 degrees Celsius. (138, 139) Following manufacturer instructions, TD tubes with brass caps and PTFE ferrules (also used by the Imperial VOC group) are validated to store VOCs for up to 27 months (Markes International, Llantrisant, UK). They should be kept in a sealed airtight container to prevent contamination,

and must be capped correctly. (140, 141) Before use, TD tubes must be cleaned or 'conditioned' by being heated to 330°C for 40 minutes in a dedicated TC20 conditioning station (Markes Ltd) following manufacturer instructions. This allows all VOCs to be desorbed away from the sorbent. A nitrogen flow of 50 ml/min is passed through the TD tubes for the full duration to ensure removal of desorbed VOCs and eliminate new contamination. After breath collection or after conditioning, brass caps must be screwed tightly onto each end of the TD tube, using a dedicated handheld spanner. This is to ensure a tight seal to prevent escape of VOCs and prevent any ambient air contamination from entering.

Advantages: TD tubes create a solution for breath storage and transport for large-scale, multi centre studies. A TD unit can be loaded with multiple TD tubes and coupled to a mass spectrometer (typically GC-MS, or PTR-ToF-MS in the case of Imperial College London), allowing automated analysis of up to 100 tubes in a run. (142)

Disadvantages: Some kind of breath collection device is required to load tubes with breath. This introduction of an extra step in the sampling process could introduce a margin of error, including the possibility of breath loss or of contamination from outside or from components of the collection system. TD tubes therefore need a QC system to ensure that sampling has occurred reliably.

3.2 **ReCIVA® breath collection system**

The Respiration Collector for *in-vitro* Analysis (ReCIVA®) breath sampling device (Owlstone Medical, Cambridge, UK) is a closed system for collecting a known volume of breath onto TD tubes. It consists of a handheld device which encompasses a sterile face mask for the patients to breathe into, where the enclosed exhalation space is emptied by the action of two electric pumps which entrain the breath across four TD tubes, at a flow rate set by the user. There is an inlet into the mask to allow control of the inhaled air. The air for inhalation is entrained via tubing into this inlet in the facemask, having been cleaned by being passing through a 'Clean Air Supply Pump for ReCIVA®' (CASPER, Owlstone Medical, Cambridge, UK). The CASPER is essentially a charcoal column acting as a filter system to absorb contaminant VOCs. See Figure 3-3 for a labelled photo of ReCIVA® being used. A standardised protocol for breath sample collection using the ReCIVA® device has been published by our group. (143)

Settings can be changed, but 500ml breath at a flow rate of 200ml/min was found to be the optimal setting, as this maximises the amount of VOCs detected but keeps the patient sampling time to within acceptable limits (approximately 5 minutes collection time). (143)

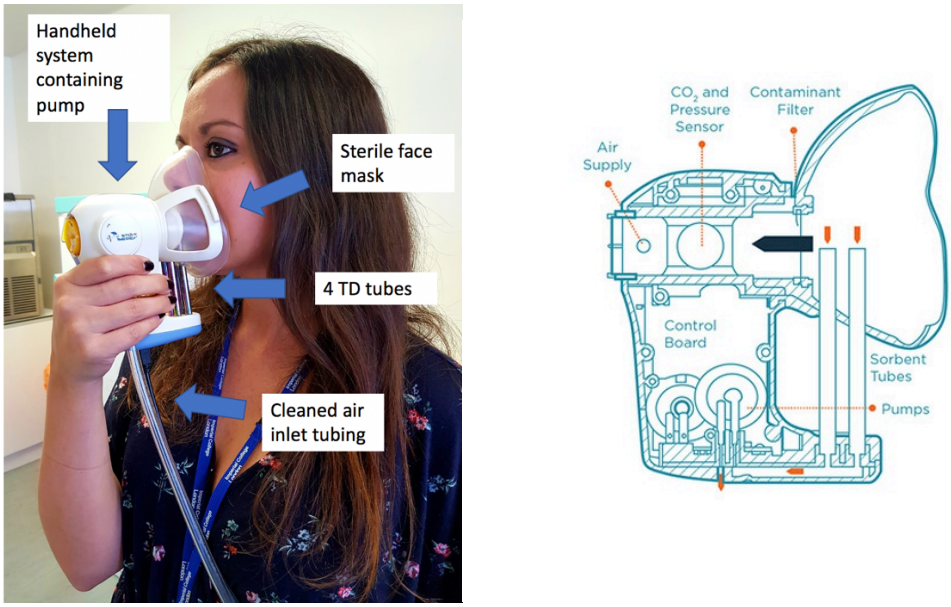


Figure 3-3 ReCIVA® breath sampler being used and diagram, adapted from the Owlstone ReCIVA® manual

A laptop controls the settings of the ReCIVA®, records quality parameters, and indicates collection completion (500ml) via a volume gauge visible on-screen. See Figure 3-4.

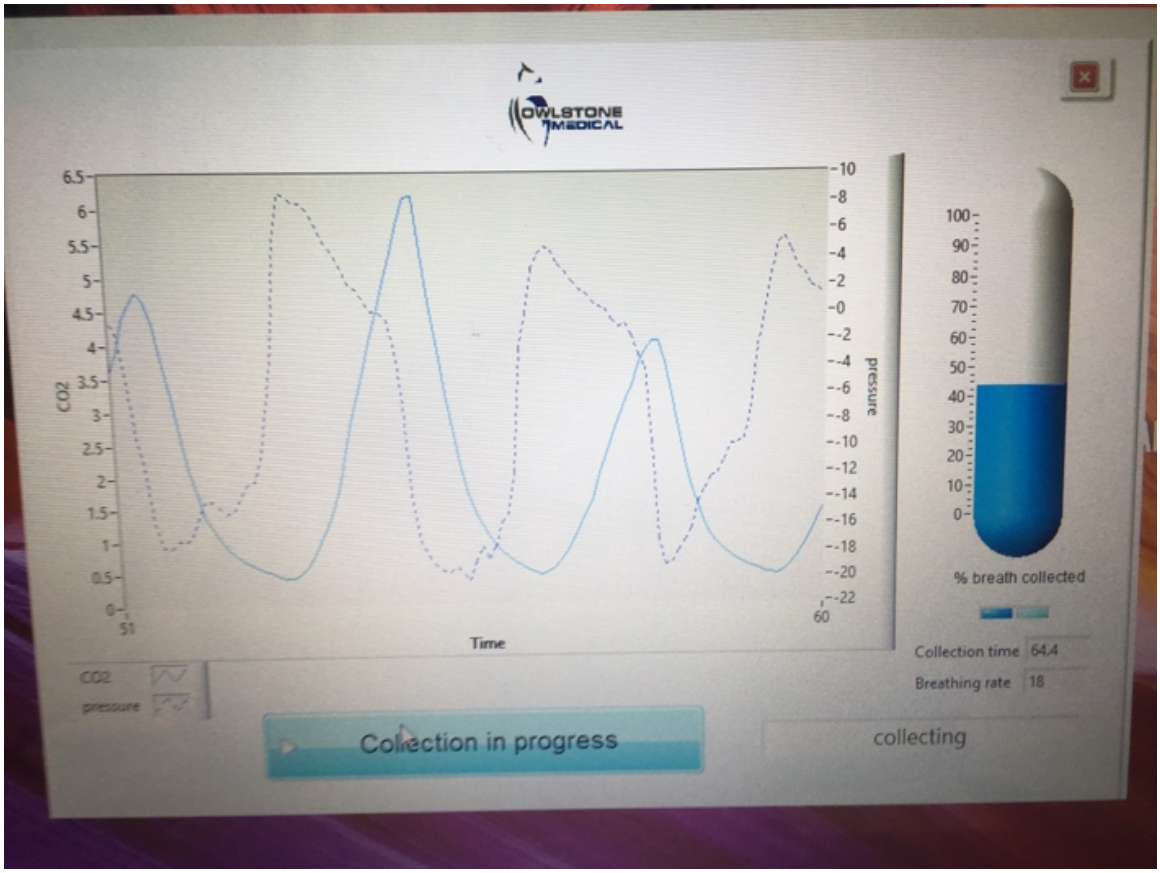


Figure 3-4 Laptop screen depicting the breath carbon dioxide and mask pressure traces during sampling

Advantages: ReCIVA® is able to preconcentrate VOCs in breath onto TD tubes by collecting VOCs from multiple exhalations. This provides a way to load breath onto TD tubes reproducibly, therefore giving all the benefits of TD tubes documented above. It is an acceptable and attractive way to sample breath (see MAGIC study in chapter 5). The ReCIVA® kit can be packed into a readily transportable suitcase, making it a realistic option for multi-site sampling. See Figure 3-5.



Figure 3-5 The ReCIVA® kit packed into a travel-size suitcase

Disadvantages: There are currently no standardised QC procedures for ReCIVA®, as it remains a minimally studied research tool. (144) A specific ReCIVA® study highlighted potential sources of error with ReCIVA® sampling, mainly regarding flow rate inconsistencies across TD tubes which affected reproducibility of VOC collection. The study concluded that to get the best and most consistent results from the ReCIVA®, the flow rates must be checked and manually calibrated, amongst other checks. (144)

3.3 Measurement of volatile organic compounds using mass spectrometry

Due to the complex composition of human breath, a number of sophisticated mass spectrometer based technologies have been employed for the detection of exhaled VOCs including the aforementioned GC-MS, PTR-ToF-MS, SIFT-MS and Ion Mobility Spectrometry (IMS). These instrumental methods all have their strengths and limitations for applications in breath analysis. All rely on the ability to ascertain the components of a mixed gaseous sample by separating ionised compounds according to their mass over charge ratio. For the purpose

of this thesis, two different mass spectrometry techniques were used to analyse VOCs in breath: GC-MS, as it is the gold standard and most widely used technique for VOC analysis in breath (82) and PTR-ToF-MS for a section of the MAGIC study, detailed in chapter 5.

3.3.1 Breath analysis using Thermal Desorption Gas Chromatography Mass Spectrometry

A TD unit coupled to GC-MS is a widely adopted method for measuring breath VOCs. Although this method requires the pre-concentration of breath volatiles prior to analysis, it allows for the comprehensive identification and semi-quantification of VOCs in unknown complex mixtures such as breath. It can separate and detect hundreds of compounds from a breath sample, allowing for VOC profiling, and has therefore been used in the study of disease biomarkers for lung cancer (145, 146), breast cancer (147), colorectal cancer (84) and liver diseases. (148, 149)

An Agilent 7890B GC-MS with 5977 A MSD (Agilent Technologies, Cheshire, UK) was used as the main analytical platform for the vast majority of the work contained within this PhD. Firstly, the components of the breath, the VOCs, had to be desorbed from the sorbent in the TD tube within which breath was collected, stored and transported. This was done by heating the TD tube to a temperature of 280 degrees Celsius for ten minutes and injecting a stream of nitrogen gas so that the VOCs passed from the TD tube to the cold trap. This was followed by secondary desorption which rapidly heated the cold trap (U-T12ME-2S; for materials emissions C4 to C32) from 10 °C to 290 °C at 99 °C/min heating rate, then held for 4 minutes to ensure complete transfer of the VOCs onto the gas chromatograph. VOCs were then separated by passage through a capillary column packed with a stationary (solid) phase, carried by inert helium gas. Separation occurred along the length the column because different molecules have different chemical properties and a different relative affinity for the stationary phase of the column. In our GC-MS the column was a ZB-624 capillary column (60 m × 0.25 mm ID × 1.40 µm d_f; Phenomenex Inc, Torrance, USA). (150) The principle was that as the components separated, they eluted from the column at different times (retention times). On leaving the column, the components were ionised by the mass spectrometer using electron ionisation sources. Ionised characteristic fragments were produced by this electron bombardment, which were then accelerated through the instrument's mass analyser (quadrupole). The electron energy applied to our system was 70 eV, the standard setting. Ions of different mass-to-charge (*m/z*) ratios deflected differently and were detected. The final

visual outcome for each separated compound was a mass spectrum, which could then be compared to known mass spectra libraries in order to identify analytes. The overall analysis was shown in a mass chromatogram, where the peak area at any given retention time corelates to the abundance of the component in question and can be compared to the same peak across different samples (although not compared to different peaks within samples). Mass spectra were recorded in full scan mode, therefore detecting all m/z ratios from 50 to 400 m/z. (150)

3.3.2 Breath analysis using Selective Ion Flow Tube Mass Spectrometry

SIFT-MS permits online, real-time VOC quantification. (136) SIFT-MS has been utilised in the study of VOCs in breath and urine from patients with conditions including cystic fibrosis and bladder cancer. (151, 152) The SIFT-MS technique allows real-time detection and quantification of VOCs within biological samples such as exhaled breath without any sample preparation, minimising diagnostic delay. SIFT-MS was not used in this PhD thesis work, mainly because it required a targeted analysis approach. Because the biomarkers for CRC and adenomatous polyps were not known, a targeted approach was not possible. SIFT-MS also needed real-time sampling or direct injection via breath bags, as our instrument was not coupled to a TD unit so TD tube use was not possible. Neither direct sampling nor bags were possible options for the large multi-centre studies of this thesis.

3.3.3 Breath analysis using Proton-Transfer-Reaction Time-of-Flight Mass Spectrometry

PTR-ToF-MS is a relatively new instrumental tool that is gaining interest in breath research applications. Similarly to SIFT-MS, PTR-ToF-MS is a direct injection-based technology, obviating problems related to pre-concentration and separation steps that are typical of GC-MS. The most important advantage guaranteed by the ToF mass analyser over quadrupole-based instruments (GC-MS and SIFT-MS) is represented by the high time resolution. In a ToF analyser a complete mass spectrum is generated at each ionisation event and spectra can be recorded at a frequency of 1 Hz or lower. Full spectral scans can also be performed without compromising sensitivity, thus enabling simultaneous targeted and untargeted analysis.

3.3.4 Sensor technology

There are many different types of sensors available, where sensors for disease detection tend to consist of an array of cross reactive sensors coupled with methods of pattern recognition. (153) Each sensor responds to a different range of VOCs, meaning that the sensors in combination are able to detect individual components from within a complex mixture of compounds. There are various different mechanisms of sensors; chemoresistors (which change their electrical resistance according to the absorption of VOCs on the sensor film) (154), acoustic sensors (which detect changes in the velocity and amplitude of acoustic waves through the sensor's coating material as a result of absorption of VOCs) (155), colorimetric sensors (where chemoresponsive dyes change colour when exposed to certain VOCs). (156) Impressive results have been achieved from sensors in detecting cancer from non-cancer patients, for multiple cancer types. (87, 117, 157, 158) Sensors are also small, often able to operate at room temperature, offer very fast results, are cheap and may have very low detection limits for VOCs of interest (sub-ppb level). (153) This contrasts to GC-MS-based techniques, which requires high levels of expertise to operate the complex and very expensive equipment, where time taken for sample analysis is much greater. (82) However, a major drawback is that the underlying identity of the VOCs that are responsible for the detection of disease groups is unknown in the vast majority of sensor-based studies. This is because generally complex mathematical algorithms (such as neural networks and machine learning algorithms) are used to build the prediction model, relying on pattern recognition developed from scrutinising a training set. The same pattern recognition algorithm is then applied to the test set, but without necessarily knowing the components of the prediction model. It is therefore difficult to be confident in a method where no mechanism can be postulated to assess plausibility and reliability. However, in the future, when selected VOCs have been proven and validated to differentiate a particular disease, specific selective sensors would likely be an excellent choice as an easy to operate, simple, cheap and crucially point-of-care device for use in primary care to detect diseases including early cancer.

3.4 **Current challenges in measuring exhaled breath**

Lack of standardisation is one of the main challenges of breath research. (159) This is the case for almost every step of the breath testing process, from equipment used, to analysis, to interpretation of results.

3.4.1 Sampling breath using the ReCIVA® breath sampler and thermal desorption tubes

Two main factors contribute to the reliable and reproducible collection of VOCs into TD tubes: stability of VOCs in TD tubes and the sampling process that transfers the VOCs there in the first place. The stability of the VOCs in sorbent tubes has been evaluated at different temperatures and over different time periods, showing an acceptable stability for approximately one month at room temperature and cold temperatures. (138, 139) The manufacturer also states that TD tubes can be stored in either the refrigerator or freezer if desired, and that for TD tubes with double sorbent (as is used in this study; carbograph/tenax sorbent), this may help "minimise risk of middle volatility analytes migrating from weaker to stronger sorbents during storage". (141) This is reassuring, but from our own tube QC checks within our lab, it is sometimes the case that TD tubes are not always completely cleaned by the conditioning process, and that VOCs on the sorbent may have persisted from a previous sample (requiring exclusion of the tube in question). The ability of the sorbent to absorb but also desorb VOCs therefore must be closely monitored, to account for inter-tube variability caused by aging of sorbent, excessive use, contamination with a particularly "sticky" VOC that is difficult to remove, or damage to a tube affecting the absorption/desorption process.

The robustness and repeatability of the sampling process using the ReCIVA® breath sampler has been recently investigated on different levels. Doran and other colleagues from the Imperial VOC group optimised sampling parameters for ReCIVA® collection; volume, flow rate and breath fraction used (whole breath/alveolar breath). (143) A recent detailed study from Harshman and colleagues highlighted interesting new findings about ReCIVA® usage and results. (144) According to this study, the best performance in terms of repeatability and accuracy of VOC measurement was obtained only with the last version of the ReCIVA® control software and by manually calibrating the flow rate at regular intervals. These findings highlight the recognised potential for variability and therefore limits of this sampling tool. Different software versions or the automated calibration of the flow rate of breath across the TD tubes can generate significant variability in the results. (144) In addition, both the studies mentioned above found a high contamination by siloxanes, which originate from the masks used for ReCIVA® sampling. However, it has been reported that the contamination from siloxanes was decreased significantly by using new batches of masks, which had been oven baked and had a slightly different design. This improvement in the manufacturing process decreased the siloxanes found. (144) However, this finding should not affect the results of the

studies in this thesis, as all cancer and control patients used the same masks, and all were sampled within the same time periods and siloxanes have been excluded from the final results. Newer masks had not been introduced when these studies were running.

These findings, supported by the poor repeatability of breath measurement in different published clinical studies (102), show the critical need in breath research to focus on many levels of QC.

3.4.2 Complexity of breath as a matrix

Another challenge is the number and variability of VOCs present in human breath, with significant inter- and intra-person variability. A study published in 2014 listed and categorised 1849 VOCs that contribute to the volatilome of a healthy human, where 874 VOCs were detected in breath. (95) Selecting breath biomarkers for disease is also complex, requiring application of advanced quantitative statistical tests, because VOCs present in cancer are mostly *also* present in healthy controls, but at different concentrations. (82)

3.4.3 Variation and lack of standardisation of analytical methods

Studies on breath VOCs have used different analytical methods, namely GC-MS (146, 160, 161), SIFT-MS (109, 162, 163) PTR-ToF-MS (161, 162, 164), and sensors (85, 87, 158), as described earlier in this chapter, making findings hard to compare. This is partly because VOC detection may vary between techniques, and VOC identification may be tentative at best or completely absent in the case of sensor technology. However, even when the most recognised and widely utilised tool is used, GC-MS, with the most established VOC identification methods using retention time and spectral library match (165) , there is an issue in comparability between results from different laboratories. (82) This is due to differences in sample collection method, sample preparation, settings of the GC-MS instrument and analysis method, GC-MS column used, and lack of validation of both the methods and results. Also, there are differences in applied calibration standards between different labs, which means that reported quantification is variable. 50% of published breath studies present no quantification of the concentration levels of potential VOC breath biomarkers. (85) Continuing improvements of qualitative and quantitative analyses have been suggested, based on the recommendations of the ISO-11843 guidelines, in order to standardise analytical procedures. (82) Future improvements would lead to better comparability of results.

3.5 Development of a quality control process for breath testing

To start to address the problem of lack of standardisation of breath testing and of the use of TD tubes, it was necessary to formalise a clear QC pathway that would address as many as possible of the multiple stages in the journey of a TD tube. An error in any one step of the process could affect quality of the breath sample and therefore integrity of the data. Steps in the pathway included transport, equipment preparation, sampling, storage and analysis; see Figure 3-6 for the different points in a TD tube's journey. A QC approach was required for the entire breath testing process, each step of which is addressed below.

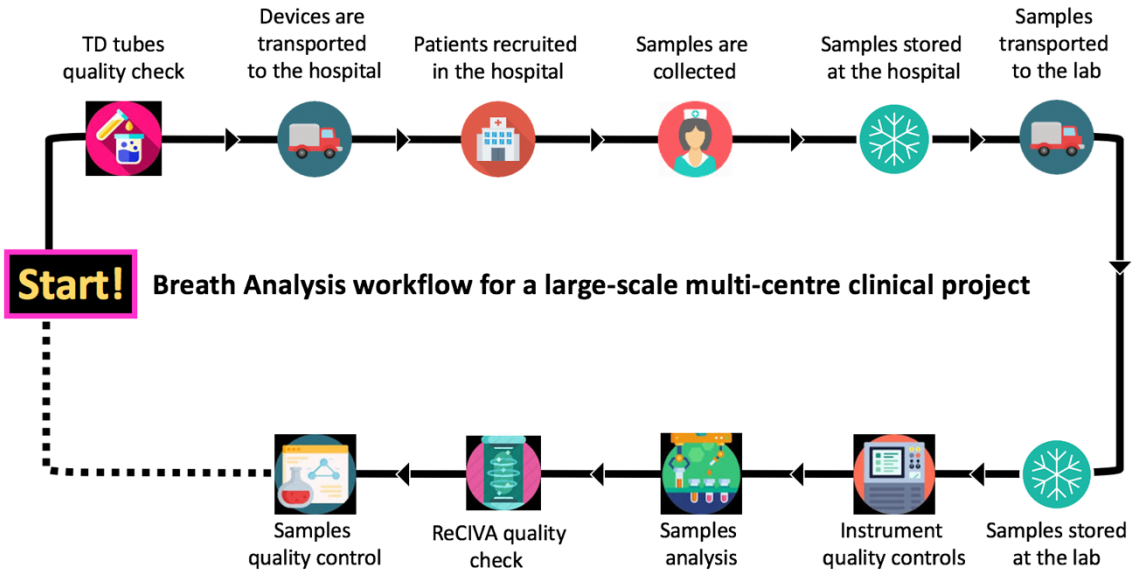


Figure 3-6 Workflow of quality control points for breath analysis

3.5.1 Thermal desorption tubes quality control

Before each use, TD tubes were cleaned with a TC20 conditioning station (Markes Ltd, Llantrisant UK) following manufacturer instructions. This essentially involved heating the tubes to 330°C whilst a flow of nitrogen was being passed through the lumen of the tube. The heat resulted in desorption of any residual VOCs within the tenax/carbograph 5TD sorbent, and the nitrogen flow ensured that no other compound could desorb from the ambient air. After 40 minutes, tubes were left to cool in a heat rack (part of the TC20 conditioning station), still under nitrogen flow, until they were cool enough to handle to allow the airtight brass caps to be replaced onto the tube ends. Brass caps were heated to 200°C in a task-specific

oven for the duration that the tubes were being conditioned. This was essentially a precaution because few VOCs would likely be emitted from the surface area of the inside of a clean brass cap.

A sample of 30 cleaned TD tubes, picked randomly from the lab stock of tubes, were tested each week to ensure efficacy of the cleaning process, as part of standard lab QC checks. This was done by evaluating the presence of known compounds on the blank tube. If any of the compounds were found to be above the 1ppb threshold, they had to be rechecked, but if any concentrations were above 1.5ppb, the TD tube had to be excluded as unacceptable. See Figure 3-7.

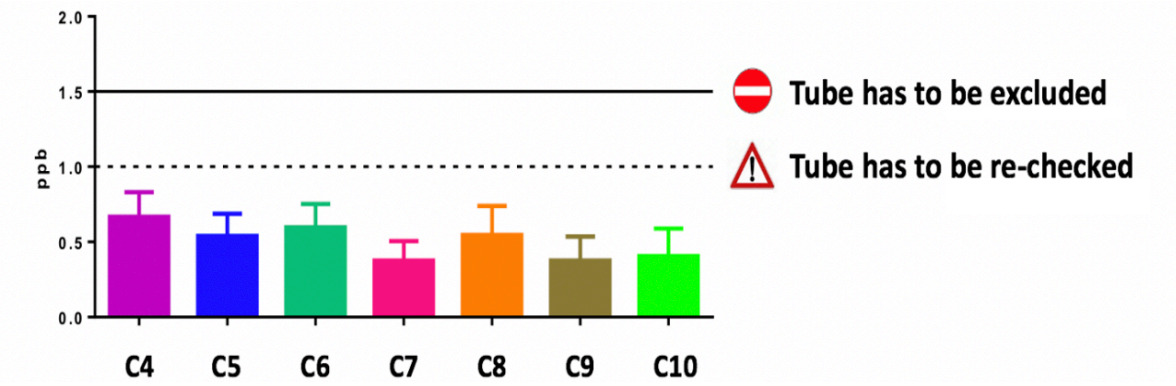


Figure 3-7 Concentrations of C4 to C10 compounds on 30 cleaned TD tubes, with quality control thresholds indicated.

As an additional QC, before any thermal desorption (either during cleaning or as part of the mass spectrometry analysis), TD tubes were logged by scanning the individual barcode of the TD tube onto a designated lab computer. A simple automated code (using the R programming language) totalled up how many times each individual tube had been desorbed over the course of its lifetime. (166) The manufacturer (Markes Ltd, Llantrisant UK) recommends that the sorbent within the TD tubes must undergo no more than 100 thermal cycles before needing replacement, to ensure high quality breath collection onto the tube. Any TD tube found to be approaching the end of its life could therefore be identified and withdrawn from general laboratory use.

3.5.2 Breath sampling equipment quality control

To ensure good functioning of the ReCIVA® breath sampler, prospective routine equipment checks were performed. This included a three-monthly check of the CASPER filtration system, the filtration column through which inhaled air was entrained before entering the sterile sampling mask. See Figure 3-8. This was done by connecting the system directly to the inlet of the PTR-ToF-MS and measuring any output of VOCs, see Figure 3-9. Figure 3-10 shows an example of a non-compliant CASPER trace, where the total concentration of VOCs which have a m/z ratio of >58 is in excess of 80ppb. Contaminated CASPERs were replaced. The manufacturer recommends replacing the CASPER after 450 hours of use, hence the rationale for the three- monthly check in our lab. The pumps in the ReCIVA® hand-held samplers were checked every four weeks, by attaching a blank TD tube to each of the tube inlets and using a flowometer to ensure that the pump was reaching the intended 200ml/min flow rate for each tube position. ReCIVA® units not achieving this were re-set and re-checked, or sent back to the manufacturer for repair. Standard operating procedure documents (user manuals) for each of these procedures, and a log book of outcomes, were kept electronically in a secure database, available to VOC group members only.

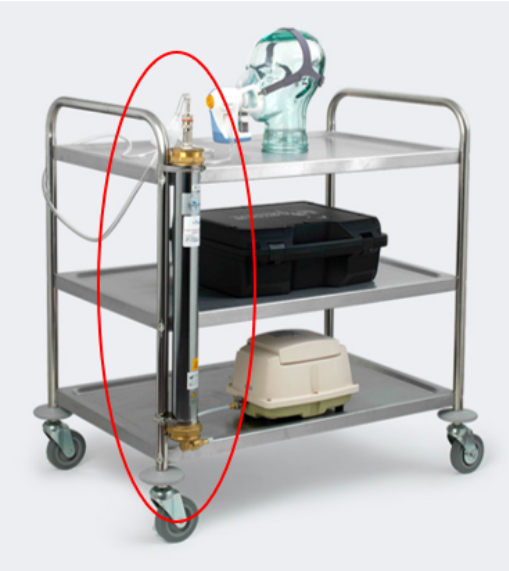


Figure 3-8 CASPER device for filtration of entrained air

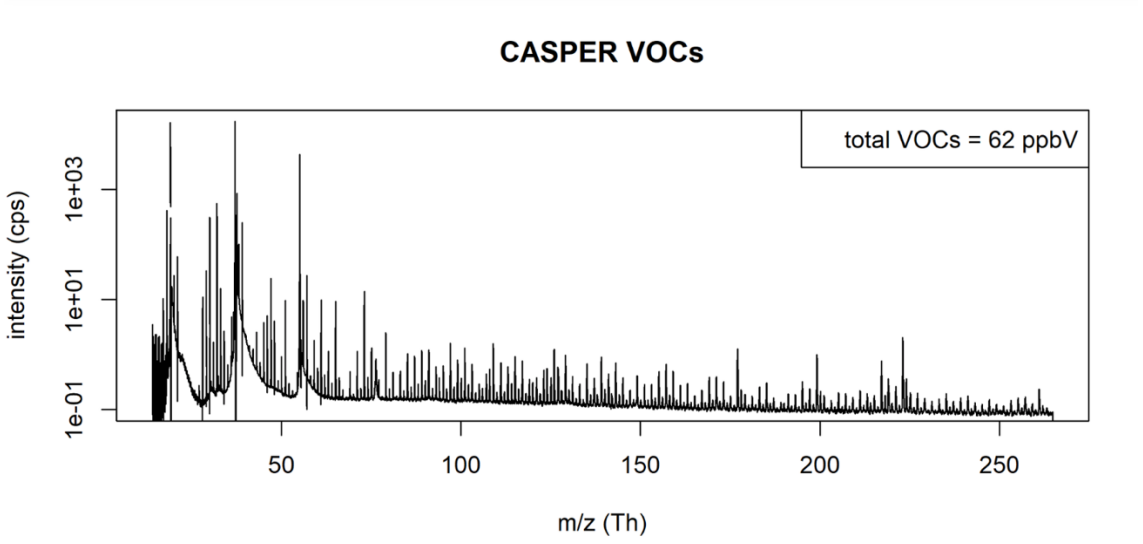


Figure 3-9 Example of a compliant CASPER VOC trace

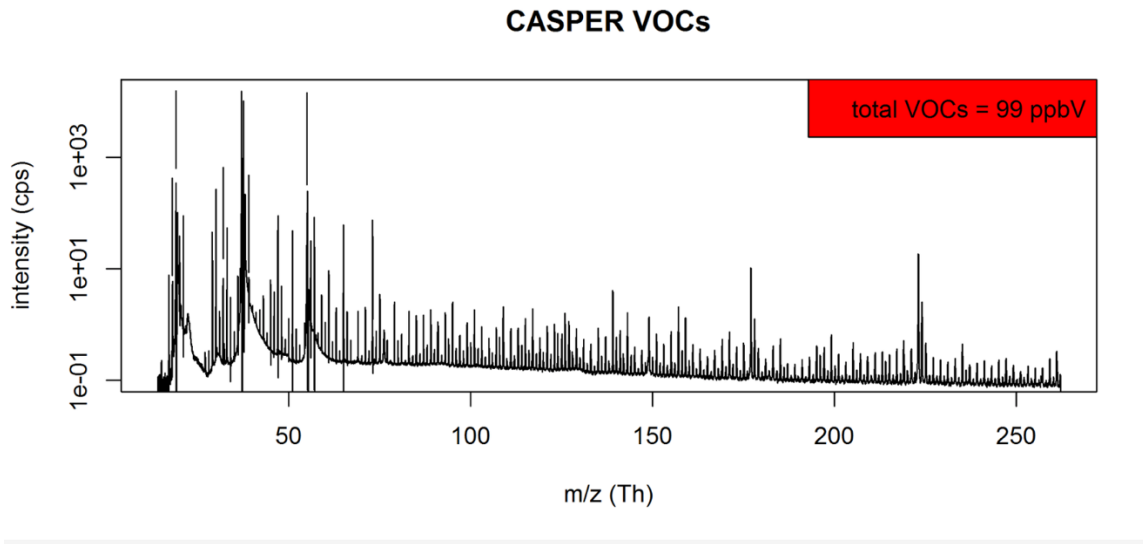


Figure 3-10 Example of a non-compliant CASPER VOC trace

In addition to these checks, research nurses were trained to monitor the volume gauge and CO_2 trace of the ReCIVA® software as the patient was breathing, in real time, during the testing. If the trace was disrupted they optimised the mask seal, or restarted the software, recording any abnormalities. If there was a software error or patient sampling interruption, which meant that the sampling ceased or was ineffective for some reason, research nurses were trained to repeat the breath sample with a new clean set of TD tubes, and label the previous failed sample as “contaminated”.

3.5.3 Retrospective ReCIVA® trace quality control checks

In addition to the pragmatic visual record of real-time sampling from the research nurses, as described above, a second retrospective check was developed to check the ReCIVA® function. Manufacturer's measured parameters were used to develop an instrumental function check for ReCIVA® using the h5 file generated by the software. This involved an in-house generated script written with R programming language (166), which interrogated each of the files created by the ReCIVA® software as each patient was sampled. Standard Operating Procedures (SOPs) for this are on the Imperial Box online database, accessible to VOC group members. Parameters recorded included the temperature of the breath, the flow rates of both of the inlet pumps (and therefore flow of breath into the TD tubes) and the pressure of the facemask against the patient's face (as a surrogate for measuring a good seal). Traces were checked retrospectively approximately every four weeks by the generation of a graphic output for each sample, see Figure 3-11 and Figure 3-12. If the parameter graphic outputs looked to have a similar visual pattern to Figure 3-11, with steadily oscillating traces as the pumps turned on and off during exhalation and inhalation respectively, then it was assumed that the software had functioned correctly. ReCIVA® traces which didn't look like this were presumed to be of poor quality due to a sampling software malfunction, and were excluded.

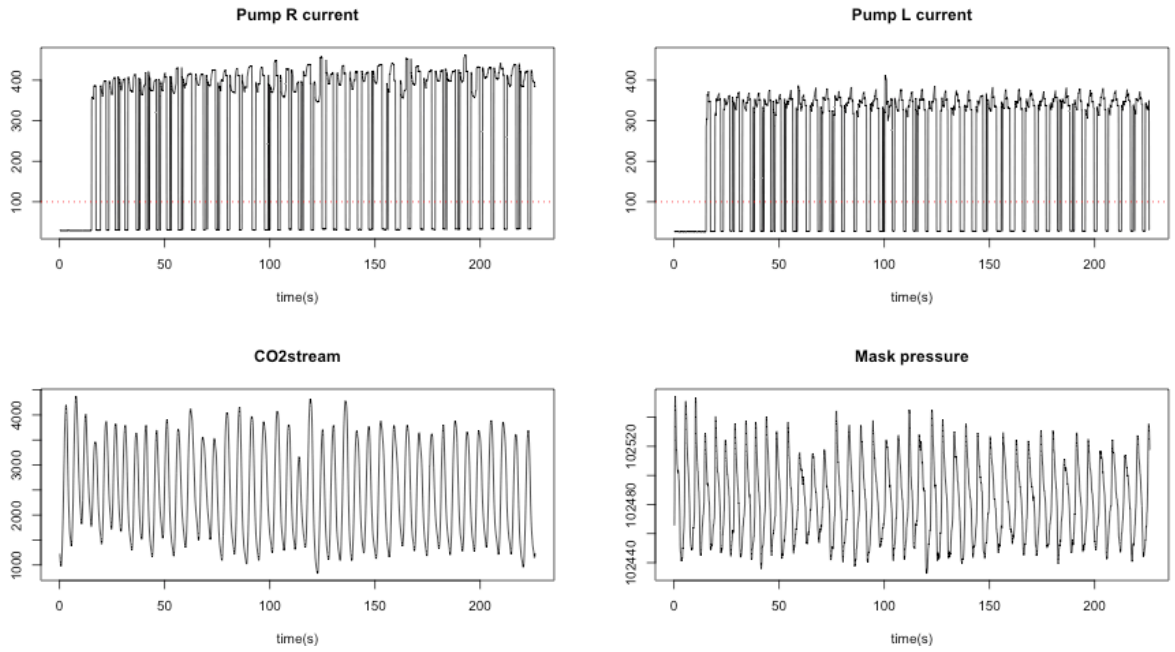


Figure 3-11 Example of the graphically represented output from a correctly performed breath sample performed with ReCIVA®

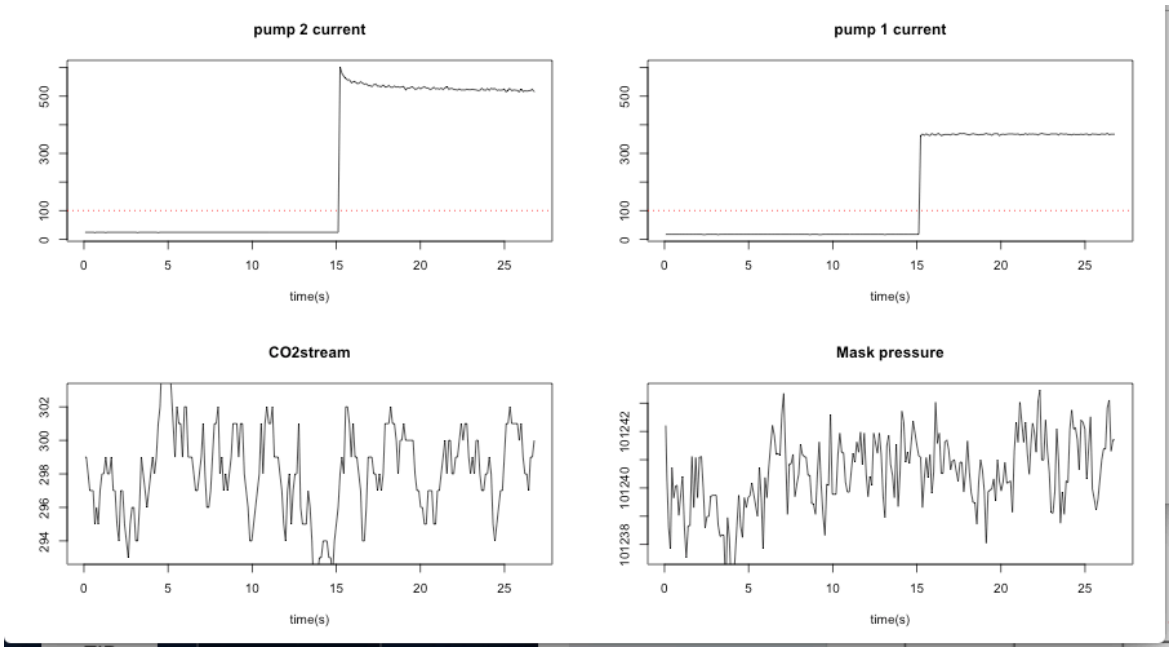


Figure 3-12 Example of the graphical representation of a poorly performed/inadequate breath sample taken with the ReCIVA®.

3.5.4 Lab instrument quality control checks

Our lab has standard QC procedures for testing efficacy and reproducibility of instruments and equipment. (142, 167)

Two types of QC were performed daily for the PTR-ToF-MS. (167) SOPs for these procedures were available to all lab users on the Imperial Box shared drive online. A first *instrument QC* evaluated instrument stability with the three ionisation modes (H_3O^+, NO^+ and O^{2+}) against four parameters: impurities, fragmentation, mass resolution and accuracy. These were measured by using a constant flow of gaseous standard VOCs with known concentrations connected directly to the PTR-ToF-MS for five minutes. This standard mixture was generated by a permeation unit (ES 4050P, Eco Scientific, Stroud, Gloucestershire UK) that provided a constant VOC concentration, maintained at 30°C and with a nitrogen flow of 0.9 L/minute. The VOCs contained in the standard mix were phenol (90 ppb), butyric acid (20 ppb), pentanoic acid (5 ppb), hexanoic acid (5 ppb), benzene (60 ppb), decanal (4 ppb) and butanal (5 ppb). Impurity levels below 10% were considered acceptable. Accuracy was evaluated through quantification of a benzene certified standard permeation tube (Kin-Tek Analytical Inc., La Marque TX). The PTR-ToF-MS quantitative measurement had to be within 20% of the certified standard in order to pass the QC. Fragmentation had to be above 60% to pass the QC. Butyric acid fragmentation was used as the check for $H3O^+$, where the ratio of diagnostic ions was used m/z 89 / (m/z 43 +71+89). For NO^+, butanal fragmentation was used, where the ratio of diagnostic ions was used m/z 71 / (m/z 43 +71+89). Resolution had to be above 1500 m/Δm. Our lab has an SOP for measuring instrument reproducibility with an action plan of what to do where parameters are not reaching the appropriate levels. No one is permitted to use the instrument for analysis until it passes the QC.

The second *standard QC* check evaluated the recovery of VOCs from TD tubes loaded from the permeation unit. (142) Tube loading was performed at a flow of 0.910 (+-0.010) L/min at a temperature of 30 degrees Celsius. This was done by connecting a pocket pump to the permeation tube inlet via a TD tube. Flow across the tube was achieved by exploiting the permeation unit flow with the addition of the pocket pump. VOCs from the permeation unit were passed through the tube for 2.5 minutes. The VOC recovery was then measured by analysing the TD tubes using the PTR-ToF-MS. This test is performed daily before lab-users can use the instrument at our centre.

For GC-MS, five TD tubes loaded with a standard mixture (as explained above) were analysed daily. Retention time, peak shape and peak area were used to assess consistency and accuracy of the instrument response.

3.6 **Creation of a sample quality control system (breath content quality control), for thermal desorption tubes**

3.6.1 Rationale

In the established QC workflow described above, there was an unaddressed step in the breath sampling process; assessment of whether breath had been absorbed into the TD tube at an adequate concentration. This was needed because it would be possible for all equipment and sampling procedures to be carried out, yet the sample still to be low quality if the sorbent had not absorbed VOCs from the breath properly. This could potentially occur if there had been an unidentified sampling or software error (e.g. correct flow rates not achieved) or if TD tubes had been handled wrongly e.g. caps not screwed on properly. Creation of a breath content QC was required to act as a final filter to try and determine whether a breath sample had been collected effectively. The presence of collected breath on TD tubes cannot be directly visualised as in other collection methods when gas inflates sampling bags or immediate results are provided by on-line mass spectrometry. (168) This means that sampling errors could potentially be missed. There are currently no operational criteria defining what constitutes a breath sample of 'acceptable quality' when TD tubes are used.

3.6.2 Aim

The aim of this study was to establish a QC method for GC-MS to evaluate the integrity of each breath sample collected onto TD tubes by determining whether or not there was indeed breath within the tube. This method was intended to be based on endogenous reference compounds that are always found in measurable quantities within human breath to serve as a fast, simple and reliable quality check for breath samples analysed by mass spectrometry.

3.6.3 Methods

Subjects between 18 and 90 years old attending St Mary's Hospital, Imperial College NHS Trust, for either routine care of a pre-existing gastrointestinal condition or follow-up for a previously treated gastrointestinal disease were recruited in the study. All patients were predominantly recruited as part of other studies run by the group, all under the overarching theme of 'non-invasive testing for gastrointestinal disease'. Ethical approval was obtained from the London - Camden & Kings Cross Research Ethics Committee (14/LO/1136). Recruited patients had consented to have their breath and clinical data used to help develop a non-invasive test aimed at diagnosing cancer of the oesophagus, stomach, colon or rectum, see Consent Form and Patient Information Sheet (appendix 3 and 4). Development of quality control procedures in order to analyse and interpret breath samples correctly was considered an instrinsic part of the breath-test development process within these studies. None of the clinical data previously collected was used for this particular QC development aspect. None were patients that were part of the COBRA or MAGIC studies, see chapter 4 and 5 of this thesis. No exclusion criteria were applied.

Sample collection

500 ml of whole breath was collected simultaneously onto four TD tubes, at a flow rate of 200 ml/min using the ReCIVA® breath-sampling device (Owlstone Medical, Cambridge, UK). A standardised protocol for breath sample collection using the ReCIVA® device has been published elsewhere, and is described in this thesis: Chapter 3, section 3.2. (143) Essentially, patients performed tidal breathing whilst wearing a sterile face mask, and their exhaled breath was entrained across four TD tubes. The TD tubes contained Tenax/Carbograph-5TD sorbent (Markes International Ltd, Llantrisant, UK). TD tubes had been cleaned in the standard fashion (described in Chapter 3, section 3.1.3) using a TC20 conditioning station (Markes Ltd) following manufacturer instructions (40 minutes at 330°C, nitrogen flow 50 ml/min). All TD tubes were sealed with brass caps and transported to the laboratory in sealed bags at room temperature, then analysed within 24 hours after collection.

A comparison control group was created by developing a group of 'poor quality' samples by mimicking errors that could happen. These were tubes that should be readily different in composition from breath samples, and should be detected as such using the new QC process. This control group was made up of (i) non-loaded (without breath on)

conditioned TD tubes; (ii) conditioned TD tubes loaded with 500 ml of room air collected using ReCIVA®; (iii) TD tubes, previously conditioned and then loaded with a standard mixture that did not contain any of the tested reference compounds. Room air for group (ii) was collected from different environments where patient sampling was performed (endoscopy admission unit and theatre admission unit). For group (iii), the standard mix was chosen based on compounds identified as possible cancer markers from previous works of the Imperial VOC group (104, 163). The standard mix was loaded on TD tubes using a permeation unit connected to the tube for 2.5minutes with a flow rate of 200ml/min, to simulate the settings used when collecting breath. The standard mixture consisted of benzene (certified standard, Kin-Tek Analytical Inc., La Marque TX), phenol (90 ppb), butyric acid (20 ppb), pentanoic acid (5 ppb), hexanoic acid (5 ppb), decanal (4 ppb) and butanal (5 ppb), generated by a permeation unit (ES 4050P, Eco Scientific, Gloucestershire UK) that provided a constant VOC concentration, maintained at 30°C and with a nitrogen flow of 0.9 L/minute.

Analysis using mass spectrometry

Sample analysis was conducted using GC-MS, described in Chapter 3, section 3.3.1. In short, a GC–MS Agilent 7890B with 5977A MSD (Agilent Technologies, Cheshire, UK) coupled with a Markes TD-100 (Markes Ltd, Llantrisant UK) TD unit was used to create a TD-GC-MS. TD tubes were desorbed using a constant flow of helium onto an electrically cooled, sorbent-packed focusing trap (U-T12ME-2S, Markes International Ltd, Llantrisant, UK), secondary desorption was then carried out by ballistic heating, leading to quick sample introduction into the TD-GC-MS. A detailed description of TD-GC-MS analytical conditions is provided elsewhere. (169) QC was performed daily for the entire duration of the project on the instrument, as described earlier in this chapter, in section 3.5.4.

Data processing and statistical analysis

GC-MS data were extracted using MassHunter software version B.07 SP1 (Agilent Technologies) and further analysis was conducted using in-house scripts developed with Python programming language (V3.7, Python software foundation). Chromatographic peak annotation was performed using the NIST mass spectral library (National Institute of Standards and Technology version 2.0). (165) The threshold level was determined by comparing the concentrations of reference compounds between TD tubes containing human

breath and those loaded with non-biological samples. VOC levels were compared using a Mann Whitney U test using GraphPad Prism (version 7.04 for Windows, GraphPad Software, La Jolla California USA), p values <0.05 were considered statistically significant.

3.6.4 Results

Collected TD tube samples with human breath and non-biological control samples are summarised in Table 3-1.

Table 3-1 Breakdown of samples collected for the breath content quality control study

Breath (biological) samples	**Control (non-biological) samples**			
Human breath samples on TD tubes (n)	Empty conditioned (blank) TD tubes	Room air on TD tubes	Standard mix on TD tubes	Total control samples
121	89	5	58	152

Acetone, isoprene, ethanol and methanol were tested as potential reference compounds, chosen as they are reported to be almost always present in human breath at measurable concentrations.

Using methanol, there was not a clear separation between breath and non-biological samples. Ethanol was not detected in some of the evaluated breath samples. It therefore could not be used as a reference standard to determine 'good quality' breath. This finding was also reported by Turner et al. (170) Isoprene was a possibility but there was an overlap between the breath and control group, making it an imperfect discriminator. Acetone (retention time = 8.97) was therefore chosen as the reference compound with a threshold of 4,000,000 peak area, corresponding to a concentration of 100 ppb. This threshold was chosen visually by seeing the ranges of concentrations within which the breath samples lay and the concentration at which the biological vs. non-biological samples separated. The threshold was kept as high as possible, with the rationale that is was better to exclude good samples than to include bad samples. Using the threshold of 4,000,000 peak area, all but five of the 121 breath samples analysed had an acetone level that exceeded this, and all 152 controls fell

below this threshold (Mann Whitney U test, P<0.0001), see Figure 3-13. Table 3-2 shows the numerical distribution of the abundance of acetone for breath and control samples.

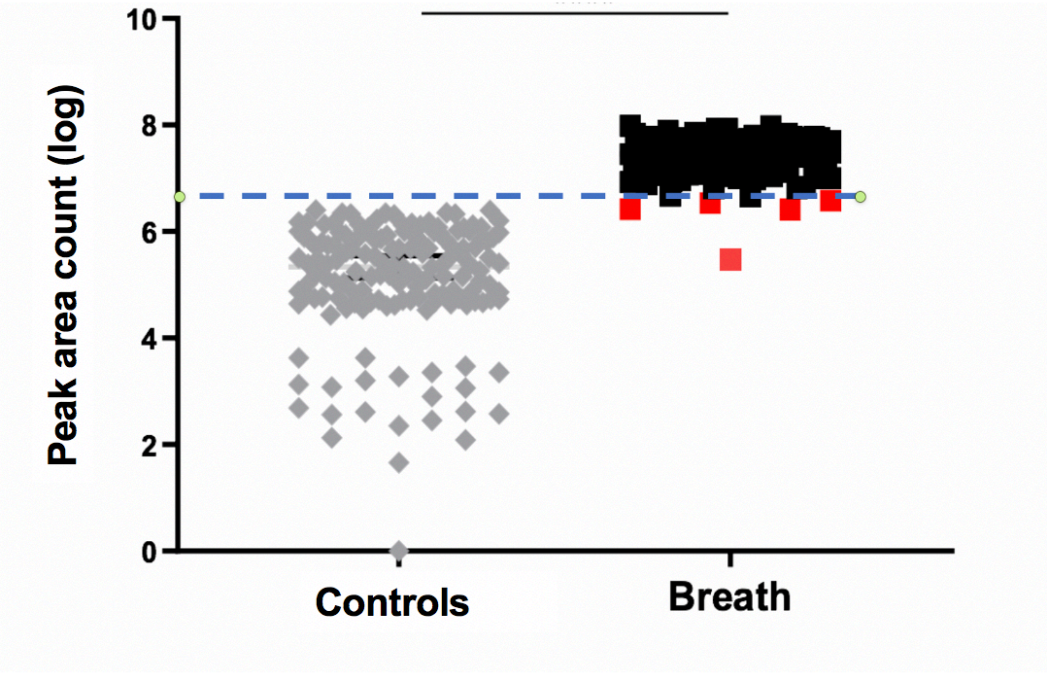


Figure 3-13 Comparison of acetone abundance in controls (non-biological) samples vs. human breath samples

The threshold of 4,000,000 area counts, (log value = 6.75), is shown by the blue hashed line. 5 breath samples that did not reach the abundance threshold are highlighted in red. All control samples fell below the postulated threshold.

Table 3-2 Acetone abundance ranges for human breath vs. non-biological control samples

Breath samples acetone abundance (peak area count)			**Control samples acetone abundance** (peak area count)			**Discriminative Threshold** (peak area count)
Minimum	Maximum	Median	Minimum	Maximum	Median	
303,448	97,047,622	28,092,455	0	2,546,486	219,358	4,000,000

3.6.5 Discussion

The results of this study provided an important contribution to the creation of a QC process for breath testing using TD tubes using GC-MS, to be used in addition to other QC steps. This was very much needed, as lack of standardisation is known to be one of the main challenges of breath research. (102, 159) A more standardised and repeatable breath testing set-up could potentially enable breath testing to become part of routine clinical practice in the future. Strengths of this study were that the threshold used peak area as the units, which meant that it could be applied to raw GC-MS data, without the need to convert to concentration in ppb. It was therefore very easy and practical to use. However, it must be noted that that this peak area threshold was specific for the given instrumental set-up and analytical method and may be affected by tuning of the ion source. If another instrument were used in a different lab, this experiment would need to be repeated to determine the optimal threshold for determining breath on a TD tube in that setting. This would also be the case for use of other analytical instruments; the same process could be developed for any analytical platform. This was in fact trialled by using the same method to develop QC thresholds for breath analysed using the PTR-ToF-MS, see chapter 5. Other limitations are that although this experiment provides a method for determining presence of breath on a tube, this does not necessarily equate to a "good quality sample". A breath sample that has been contaminated by environmental factors such as contaminated room air could plausibly pass the QC on the basis of the level of acetone, but not be a representative breath sample. In the same argument, breath could be lost/VOCs desorb from the sorbent but as long as the

TD tube reaches the particular acetone threshold it would still pass the QC (for example if a patient had a particularly high acetone level from a prolonged fast, some VOC loss would be undetected by this method); it is therefore a rough guide only.

3.6.6 Conclusion

In this study, a QC method was established to assess the presence of breath in TD tubes that could serve as a simple and effective standardisation parameter for use in multi-centre clinical trials. This level of QC was required because there was no way of confirming visually the presence of breath on a TD tube, where inadequate sampling may lead to inaccurate results.

3.7 **Summary of chapter**

This chapter introduced the multiple sampling and analytical methods used for breath VOC research in the literature, and explained the rationale for the use in this thesis of the ReCIVA® with TD tubes, analysed by GC-MS. The assembly of a QC work flow with the new addition of breath content QC improved the breath analysis process in our lab by i) creating an established structure; ii) increasing awareness of QC at every stage of the process; and crucially iii) making the process more robust and repeatable. This QC process became standard practice in our lab, and as we continue to learn the lessons of putting it to the test in large clinical studies, results will be disseminated to the wider breath community, and quality can continue to improve.

4 CHAPTER 4: PROFILING OF VOLATILE ORGANIC COMPOUNDS IN BREATH FOR THE DETECTION OF COLORECTAL CANCER AND ADENOMATOUS POLYPS, THE COLORECTAL BREATH ANALYSIS (COBRA) STUDY

ClinicalTrials.gov Identifier: NCT03699163

4.1 **Study Aim**

Investigate the use of volatile organic compounds (VOCs) present in exhaled breath for the prediction of colorectal cancer (CRC) and adenomatous polyps.

4.1.1 Objectives

- Collect and compare breath VOCs from a large cohort of patients with CRC, adenomatous polyps, benign diseases of the colon, and no colonic disease, as diagnosed on colonoscopy;
- Investigate the diagnostic accuracy of the breath test in a group of patients who have CRC and adenomatous polyps compared to subjects with benign diseases or normal colons, by constructing a diagnostic model;
- Identify and biologically characterise any significant ions.

4.2 **Ethics**

The COBRA study was given research ethics committee (REC) approval on 28/4/17 (17EE0112), and Health Research Authority approval on 2/5/17 (East of England- Essex REC). Site Specific Assessment was also carried out at all 7 participating hospitals, with approval of the study sponsor.

In order to sample breath from patients enrolled into the English BCSP, COBRA received specific approval from the BCSP Research Advisory committee (BCSP ID189, approval given on 18/1/17). In addition, COBRA was adopted into the NIHR portfolio. This allowed recruitment to be conducted by NIHR affiliated research nurses.

The study was conducted in accordance with the recommendations for physicians involved in research on human subjects adopted by the 18th World Medical Assembly, Helsinki 1964 and later revisions.

4.3 **Methodology**

COBRA was a prospective, non-randomised, cohort study designed to sample the breath of patients having colorectal investigations in secondary care at 7 London hospitals, over 3 years, starting on 5th June 2017.

4.3.1 Inclusion Criteria:

- Participants between the ages of 18 years and 90 years inclusive.
- Able to provide informed written consent.

  AND
- Undergoing a lower gastrointestinal endoscopy (colonoscopy) as part of their routine clinical care.

  OR
- Scheduled to undergo elective resection of histologically confirmed colorectal adenocarcinoma.

4.3.2 Exclusion Criteria:

- Lacked capacity or unable to provide informed consent.
- Any patient below 18 years of age or over 90 years of age.

4.3.3 Patient selection - endoscopy unit

Patients were invited to participate in the study whilst waiting for a planned colonoscopy in the endoscopy unit of one of 4 participating London-based BCSP endoscopy centres: Charing Cross Hospital, St George's Hospital, St Mark's Hospital or Homerton University hospital. Patients attending for a BCSP colonoscopy were approached preferentially because their chance of having a colonic polyp was estimated to be around 40% (48) and chance of CRC higher than in the general population (given that all BCSP attendees were by definition faecal occult blood test-positive at the time of sampling). However, any other patients attending for a colonoscopy were also eligible, including those attending for 2WW or surveillance colonoscopies. Patients sampled in the endoscopy unit were not pre-selected before the day, they were sampled as they came, and neither the study organisers nor the breath samplers had seen any of the patient's medical history or records prior to sampling them. All had been referred for a colonoscopy either on clinical grounds by their usual medical practitioner, or were attending as part of the BCSP. All patients were nil by mouth and had fasted for a minimum of 6 hours as per usual endoscopy guidelines. Patients were sampled in a side room away from other patients, in a seated position, before entering

the endoscopy procedure room. This was done to avoid any effects of sedative drugs or anaesthetic throat spray present in the endoscopy room itself.

4.3.4 Patient selection - theatres

An additional cohort of patients was approached for the study who were known to have current active CRC, specifically colorectal adenocarcinoma in situ (within the colon), with a planned upcoming surgical resection of the tumour. These patients were identified by the study team by reviewing the lists from the weekly hospital multidisciplinary team meetings for CRC patients, in one of 3 participating London hospitals; St Mary's hospital, West Middlesex hospital or Chelsea and Westminster hospital. Once patients had been identified, their online medical records were checked by the local study investigator. Included patients were not taking chemotherapy at the time of the operation. Patients were approached on the morning of their surgery to ask if they would give a breath sample before their operation. All patients were nil by mouth and had fasted for a minimum of 6 hours as per usual theatre guidelines. Patients were sampled in a side room in the surgical department away from other patients and separated from theatres, in a seated position. All breath samples were retrieved prior to the anaesthetic or surgical procedure, before transfer to the anaesthetic room.

4.3.5 Target study size calculation

The study target was 1463 patients. This study size target was based on a power calculation showing that if CRC has an 8% prevalence in the intended cohort and the test has an expected sensitivity of 85%, 117 cancer patients would be needed and 1346 controls, a total of 1463. Prevalence of CRC was estimated at 8% for CRC, and 40% for adenomatous polyps. 40% polyp prevalence is consistent with BCSP literature, however a lower prevalence of CRC was chosen than the reported 10% (48), because COBRA was planned as a prospective study and the incidence of CRC is lower in London in comparison to the rest of England; incidence for males in London was 70.5 per 100,000 compared to 95.1 per 100,000 in the North East of England, from the National Office for Statistics, 2017. (2) The sensitivity estimate of 85% was based on the assumption that this breath testing study would outperform previous results obtained from our group (sensitivity for oesophageal cancer found to be 80%, sensitivity for CRC found to be 83%). (109, 163) This assumption was based on the fact that

this study was planned to use a novel mass spectrometry multi-platform approach and a higher number of recruited patients.

In addition, we aimed to over-recruit by 15%, to account for inevitable data loss. Potential reasons for this could be instrumental failure, human errors with sample analysis, or failure of samples to reach desired QC standards. (144) Accounting for a 15% data loss, we therefore aimed to recruit 1722 patients.

4.3.6 Patient sampling

Patients were sampled in an identical fashion regardless of whether they were recruited from endoscopy or theatres. Patients were approached by a member of the research team to discuss the research project and be given a Participant Information Leaflet (see appendix 5). If they agreed to take part, a written consent form was signed (see Consent Form - appendix 6) consenting to donate a breath sample and allow their clinical details to be recorded.

The breath test involved participants performing normal tidal breathing whilst wearing a sterile rubber facemask (single use) fitted onto the ReCIVA^{TM} CE-marked handheld breath testing device (Owlstone, Medical Ltd, Cambridge, UK), as per published optimised settings. (143) The procedure and settings for breath testing are described in chapter 3 of this thesis. In brief, during exhalation breath was entrained from the mask via four TD tubes (Markes International, Llantrisant, UK) at a flow of 200mls/minute using inbuilt pumps (triggered by rising carbon dioxide levels), having a final volume of 500 ml per tube. The TD tubes were packed with Carbograph/Tenax sorbent phase, designed to retain VOCs. The 'whole breath' setting for breath fraction was chosen. After the breath test (which lasted approximately 5 minutes), a member of the research team was responsible for sealing the TD tubes by screwing brass caps onto each end with a specific spanner, to ensure that the breath VOCs were trapped onto the sorbent in the TD tube and could not desorb and escape. Researchers also filled out a clinical details form, detailing past medical history, BMI, medications and key information such as smoking status and last meal (see Baseline clinical data form - appendix 7). Sets of four capped TD tubes were then placed in plastic sealed sampling bags, labelled with the unique study identifier, date and time and site of sampling. Operators wore non-sterile hospital gloves throughout. The used mask was disposed of.

4.3.7 Breath sample handling

The samples collected and clinical details forms were labelled with the same unique study identifier (anonymised). The sample bags containing capped TD tubes were kept in a sealed airtight Tupperware container. At the end of the week, or after the endoscopy or theatre list (depending on volume of samples collected at each site), samples were collected by a medical courier or by a member of the research team to be taken to the central VOC laboratory at St Mary's Hospital London. Breath VOCs sampled onto TD tubes were analysed usually within 12 hours from arrival. If a delay of more than 48 hours was anticipated before analysis, then the TD tubes were placed in the -80^{o} C freezer, still in airtight Tupperware containers. Analysis start time (and freezing time if relevant) were recorded onto a database.

4.3.8 Specimen analysis

Breath VOCs were analysed using two mass spectrometry techniques previously described: PTR-ToF-MS and GC-MS, introduced in chapter 3 section 3.3.

Three of the four TD tubes from each patient were analysed using PTR-ToF-MS (using three different reagent ions H_3O^+, NO^+, O_2^+), and one TD tube using the GC-MS. The GC–MS Agilent 7890B GC with 5977A MSD (Agilent Technologies, Cheshire, UK) was used, coupled with a Markes TD-100 (Markes Ltd, Llantrisant UK) TD unit. The analytical method is described elsewhere. (169) GC-MS analysis was performed with a two-stage desorption method using a constant flow of helium at 50 ml/min and a cold trap system (U-T12ME-2S, Markes International Ltd, Llantrisant, UK). Samples were then transferred to the GC system by a capillary heated at 200°C. The chromatographic column employed for compound separation was a Zebron ZB-642 capillary column (60m x 0.25mm ID x 1.40 µm df; Phenomenex Inc, Torrance, USA).

GC-MS data were extracted using MassHunter software version B.07 SP1 (Agilent Technologies) and further analysis was conducted using a custom designed in-house built software MSHub. (171, 172) VOC peak identification was performed using the NIST mass spectral library (National Institute of Standards and Technology version 2.0). (165)

4.3.9 Rationale for measurement using different analytical platforms

As described in the previous chapters, including in the systematic review of the literature, GC-MS is considered the gold standard for the analysis of VOCs in breath. For this

reason, we chose to use this platform, characterised by high reliability and good VOC identification performance. PTR-ToF-MS is a novel technique, used in environmental research. PTR-ToF-MS is characterised by high-throughput and real-time results. In contrast to GC-MS, PTR-ToF-MS provides direct quantification of compounds, without the need for external calibration. These aspects make the use of the two techniques complementary. GC-MS offers reliable compound identification while PTR-ToF-MS offers high-throughput analysis and quantitative results. For this reason, GC-MS was used as a "discovery" technique, while PTR-ToF-MS was used to provide a fast real-time method. For the biomarker identification purposes, only GC-MS data will be discussed. The ReCIVA® breath sampler has the ability to collect four breath samples simultaneously, allowing two mass spectrometry platforms to be used without adding additional breath sampling time for the patients.

4.3.10 Staff and Training

At each site there was a team responsible for identifying patients (in the case of surgical patients), approaching patients, consenting, sampling breath, filling out clinical details forms, managing samples and coordinating with the central study team for sample collection. These staff were research nurses and research assistants from the local hospital research and development departments, under the auspices of the NIHR clinical research network. In one case (St George's Hospital, Tooting), clinical research nurses were also from the local Clinical Research Facility, as well as specialist endoscopy nurses who had been trained to breath-test. Human factors analysis had already been carried out for other VOC studies from Imperial College, demonstrating several potential sources of human factor related error in breath sampling that could affect the results of an analysis. (163) All staff members were therefore trained on site in the environment that they would be sampling in, 1-3 weeks before recruitment opened at the site, over approximately 2 hours. This included observation of the breath-testing process. An instruction booklet was provided with each ReCIVA® portable device, with a step-by-step guide of how to set up the equipment. This was also emailed to all research nurses prior to the start of the study (see ReCIVA® List of tasks - appendix 8). This ReCIVA® manual was developed and tested by colleagues who had never set up or used the ReCIVA® before using the Think Aloud Protocol. (173) All staff also had a copy of the study protocol (see appendix 9).

4.3.11 Quality Control

QC was carried out at each of the following stages:

- TD tubes QC (conditioning of TD tubes and weekly testing of blank tubes)
- Sampling QC (real-time monitoring of sampling process)
- Instrumental lab checks (instrument QC and standard QC)
- ReCIVA® trace QC checks
- Breath sample QC (breath content check)

These stages are all described in detail in chapter 3.

4.3.12 Data analysis

Demographics and clinical data

Patients were grouped according to the findings on the colonoscopy/surgical procedure that they had on the day of performing the breath test. The normal group (group 1) were defined as those with no abnormal findings on colonoscopy . The benign pathology group (group 2) were defined as those minor non-inflammatory findings; haemorrhoids, benign non inflammatory anal fissures, diverticular disease or benign diverticular strictures. The IBD group (group 3) was defined as those with colonoscopic evidence of UC, CD, unspecified colitis or infective colitis, of any severity, corroborated by histology results from the colonoscopy. If patients had a history of IBD documented in their records, but had a normal colonoscopy with normal biopsies, these patients were allocated to the 'normal' group. Those with adenomatous polyps were stratified in to low risk (group 4), intermediate risk (group 5) and high risk (group 6) categories based on the theoretical risk of polyp development into CRC. This was calculated using adapted criteria taken from the British Society of Gastroenterology polyp surveillance guidelines 2002, adapted to include dysplasia status, and the more recent guidance on sessile serrated polyps from 2017: (8, 174)

Low risk polyp patients:

- 1–2, small (<1 cm) tubular adenomas with low grade dysplasia

Or

Sessile serrated polyps (SSPs) <1cm with no dysplasia (incudes hyperplastic polyps)

Intermediate risk polyp patients:

- 3–4 small tubular adenomas, with low grade dysplasia.

Or

- At least one adenoma>1 cm, with low grade dysplasia.

Or

- SSPs >1cm, with no dysplasia

High risk polyp patients:

- ≥5 adenomas

Or

- ≥3 adenomas where at least one ≥1 cm.

Or

- Any adenoma with high grade dysplasia

Or

- Any adenoma with any villous change (including tubulovillous adenomas)

Or

- Any SSP with evidence of dysplasia

CRC patients (group 7) all had colorectal adenocarcinomas, where size, site, grade of tumour and TNM stage were documented. Polyposis patients (group 8) were those with an existing diagnosis of polyposis, corroborated by colonoscopic findings (polyposis diagnoses included familial adenomatous polyposis (FAP) where colectomy had been refused, serrated polyposis, Lynch syndrome, juvenile polyposis and MUTYH associated polyposis). Colonoscopic diagnosis was determined as per the most significant finding. The diagnostic group hierarchy was CRC, polyposis, high risk polyp(s), intermediate risk polyp(s), low risk polyp(s), IBD, benign pathology, normal. This meant that if a patient had IBD and a polyp, regardless of whether it was active IBD or not, they were placed in the appropriate polyp group. In the same way, a high risk polyp categorised patient could also have a diverticulum or haemorrhoid.

Demographics for all patients were recorded, including tumour-specific factors where relevant. Potential confounding factors across the CRC and control groups were evaluated using the Mann-Witney U test for continuous variables and χ^2 test for discrete variables. $P < 0.05$ was used to assign statistical significance. This statistical analysis was performed using the statistical software SPSS (version 25, IBM).

Breath VOC data

The raw data from the GC-MS analysis were processed with MSHub, a custom-made spectrum processing program, made at Imperial College London. (171, 172) This was a dataset-based spectral deconvolution tool for use within the Global Natural Product Social Molecular Networking (GNPS) environment. The steps by which MSHub processed the raw data were: intra/inter-sample mass drift correction, noise filtering and baseline correction, inter-sample peak alignment, peak detection and integration, non-negative matrix factorisation (NMF) deconvolution then peak deconvolution. (171, 172) This gave an output that consisted of multiple ions (or VOCs) labelled as numbered features, their retention times, and the peak area count of each feature in each patient's breath sample. Not all features were present in all samples. In addition, there were some features identified which were ions that made up a very small proportion of the total peak for a given retention time, present in the minority of samples. Ions that made up less than 20% of the total peak were included in the statistical analysis, but were not considered in the list of top differentiating features in the comparisons of different clinical patient groups. The target ions from the obtained spectra were matched using the on-line NIST library for potential identification. (165) The MSHub utilises a one-layer neural network for GC deconvolution, which allows information to be extracted across the entire dataset (as opposed to a single spectrum at a time) and thus utilises all of the spectral information within the data, a strategy that is particularly successful for large-scale studies.

Statistical analysis

Both univariate and multivariate data analysis techniques were applied to the results in order to:

1) Identify VOC components with the best discriminating ability between the groups
2) To develop a multivariate discriminant analysis model.

A Mann-Whitney U test was used to compare the measured VOC levels between selected groups, namely CRC vs non-CRC groups, or to investigate potential confounding factors such as sampling environment, tube storage times, or anatomical site of tumours. A p value <0.05 was taken as the level to indicate statistical significance. A non-parametric (Kruskal-Wallis) ANOVA test was used to compare the measured VOC levels (VOCs represented as ions) between all 7 of the included study pathology groups (grouped according to diagnosis as per

colonoscopy result). This was done to determine if any of the 7 patient groups contained an abundance of a VOC that was statistically significant in differentiating between groups. A p value <0.05 was taken as the level to indicate statistical significance. This basic statistical analysis was performed using the statistical software SPSS (version 25, IBM). (175)

Clinical parameters that required further investigation to establish any confounding influences on VOC abundance, such as fasting time and T stage of the tumour, were investigated with Pearson's correlation coefficient (in the case of fasting time), and by plotting VOC abundance trends in the case of tumour T-stage comparisons. This was done using SPSS (version 25, IBM) and Microsoft Excel v16.43.

Machine learning prediction models

A high performance computer facility at Imperial College London was utilised to run a machine learning pipeline to process all of the abundance data of unidentified features in each patient's breath sample (1024 features were identified in each sample), and the extensive metadata for each patient. The data was normalised, variance stabilised and log-transformed as part of the machine learning pipeline. Random forest, alphanet, SVM, lasso and elastic machine learning prediction methods were used independently to compare every combination and permutation of pathology group. The same analyses were repeated also for patients of age 40-59 years, 45-65 years, 50-69 years and 70-89 years, as well as all ages together, to investigate whether age was confounding the VOC data. The prediction models took into account a wide range of clinical variables between groups. These included patient factors: age, number of hours of fasting, BMI, ethnic origin, gender, smoking status, weekly alcohol consumption, type of bowel preparation taken before colonoscopy/surgical resection and family history of CRC. Sampling related factors were also included: the method by which TD tubes had been cleaned before sampling (using the standard TC20 conditioning unit as explained in chapter 3, or using the PTR-ToF-MS instrument itself), the storage time of the TD tube from conditioning to breath sampling, the storage time post-sampling until mass spectrometry analysis, and the number of days the TD tube was stored in the freezer (if applicable). Factors that were directly linked to outcome were excluded from the prediction model, such as reason for colonoscopy, sampling site, and any data linked to colonoscopy findings. Details of past medical history and medications were not inputs into the model, as answers were too numerous and heterogeneous. This would have been a problem because if

every answer was coded, “missing data” would have been very common and would have introduced bias, whilst only sporadic mention of a variable would have added unnecessary noise to the dataset and not have been useful.

ROC curves were used to determine the accuracy of a diagnostic test in classifying those with and without colorectal disease. The ROC curves were generated based on 25 runs: 5 repeats of 5-fold stratified K-fold splits with re-shuffling between splits. This meant that samples were shuffled and then split into 5 groups. Each group was then used in turn as a test set, while the other 4 were the training set. Feature selection and model building (machine learning) were performed on a training set each time (80% of the data) and then applied to the test set (20% of the data) to produce the statistics. This was repeated 5 times and then the results from different runs were averaged to get ROC curves and error estimates. This way overfitting of the model was avoided (as if we were to train and test on the same data, we could have easily achieved high prediction levels, but then when any new future data came, the prediction would fail). Also it avoided the bias of an “unlucky” split if the data was randomly split in a non-representative way. Because this analysis method was chosen, each time the data was split, the selection of significant features varied slightly.

The average number of times any given feature was selected as a predictive/significant feature was displayed as a feature selection score. If a feature was independently selected to be a differentiating feature regardless of how the data was split, the selection score would be higher. A higher score therefore meant that the feature in question was more likely to be a true feature differentiating CRC and non-CRC, as opposed to a chance finding.

In addition, in the case of Random Forest (RF) method, the contribution that each feature made to the prediction model was represented by the RF score. The scores for all features contributing to the generation of the predictive model always added up to 1 (by definition). The highest scoring features therefore represented the most important in terms of differentiating the comparator groups. The score was calculated by computing the normalised total reduction of the criterion brought about by that feature (also known as the Gini importance). (176)

4.4 **Results**

4.4.1 Recruitment

1855 patients were recruited between 5th June 2017 and 21st February 2020, over 7 NHS hospitals in London. Centres commenced sampling at staggered intervals as the study was set up, and terminated as individualised recruitment targets were reached per site. See Table 4-1 for details.

Table 4-1 Breakdown of COBRA recruitment by NHS centre

Site	Date commenced	Date ended	Sampling period (days)	Number from endoscopy	Number from theatre	TOTAL FROM SITE	% from site
Charing Cross Hospital	05/06/2017	17/04/2019	681	520	0	520	28%
St Mary's Hospital	05/06/2017	19/08/2019	805	1	81	82	4%
St Mark's Hospital	28/06/2017	05/04/2019	646	243	0	243	13%
St George's Hospital	26/09/2017	26/03/2019	546	704	0	704	38%
Homerton Hospital	11/01/2018	01/04/2019	427	249	0	249	13%
West Middlesex	26/05/2018	21/02/2020	636	0	27	27	1%
Chelsea and Westminster	03/07/2018	23/10/2019	477	0	30	30	2%
TOTAL PATIENTS RECRUITED				**1717**	**138**	**1855**	

Declining to participate

Recruitment figures above are from those who agreed to take part in the COBRA study and who gave a breath sample. More patients than this were approached, but declined to take part so were not included in recruitment figures. See the total approached patients and decline rate in Table 4-2. Reasons for declining included language barriers (where interpreters were not available), patients not having their glasses to read the consent form, unwillingness to participate due to anxiety about their upcoming procedure, and anxiety about breathing into a mask. Also recorded in Table 4-2 below are patients that were approached about the breath test, and consented to participate, but then the breath sample was never collected

"pre-test exclusions". These patients could not be included in recruitment numbers as they had not been sampled, but also could not be counted within the declined numbers. This occurred largely because the sampling nurses judged that there would not be enough time to do the breath sample before the patient's procedure, so explained to the patient that this must take precedence over the research study. The study team were trained to have a low threshold for doing this, to ensure that no patient had the flow of their planned care disrupted.

Table 4-2 Patient approach and decline rate, and pre-test exclusion rate per site for COBRA study

Site	Total approached	Declined	% decline rate	Patient agreed but breath sample not collected (pre-test exclusions)	Total successful patient recruitment	% Successful recruitment (of all approached)
Charing Cross	678	8	1%	150	520	77%
St Mary's Hospital	94	4	4%	8	82	87%
St Mark's Hospital	286	3	1%	40	243	85%
St George's Hospital	873	44	5%	125	704	81%
Homerton Hospital	331	1	0%	81	249	75%
West Middlesex	31	4	13%	0	27	87%
Chelsea and Westminster	41	9	22%	2	30	73%
TOTAL	**2334**	**73**		**406**	**1855**	

Exclusions and lost data

Of the 1855 eligible patients who had given breath samples, a significant proportion, 22.8% (n=423), were not included in the final analysis. Figure 4-1 details the reasons for all exclusions/data loss. In summary, whilst some data loss were beyond anyone's control (incomplete/cancelled/declined colonoscopy post breath test (n=44)) and some data loss was due to simple human error in patient identification/sampling settings/sample handling/ analysis (n= 68), the largest reason for exclusion was on poor QC grounds. Failure to take a good quality sample made up 54% of the cause of data exclusion (61 from ReCIVA software

errors or exclusion post ReCIVA QC, 166 from failure of breath content QC using acetone thresholds-assumed to be due to poor sampling). Instrumental analysis faults due to instrumental breakdown/file corruption was the next highest cause for data loss, 68 samples, accounting for 16% of the data lost, equal to the data loss caused by human errors. See Figure 4-1 for the detailed breakdown of the exclusions leading to the final included patient number of 1432. This flow chart focuses on GC-MS data.

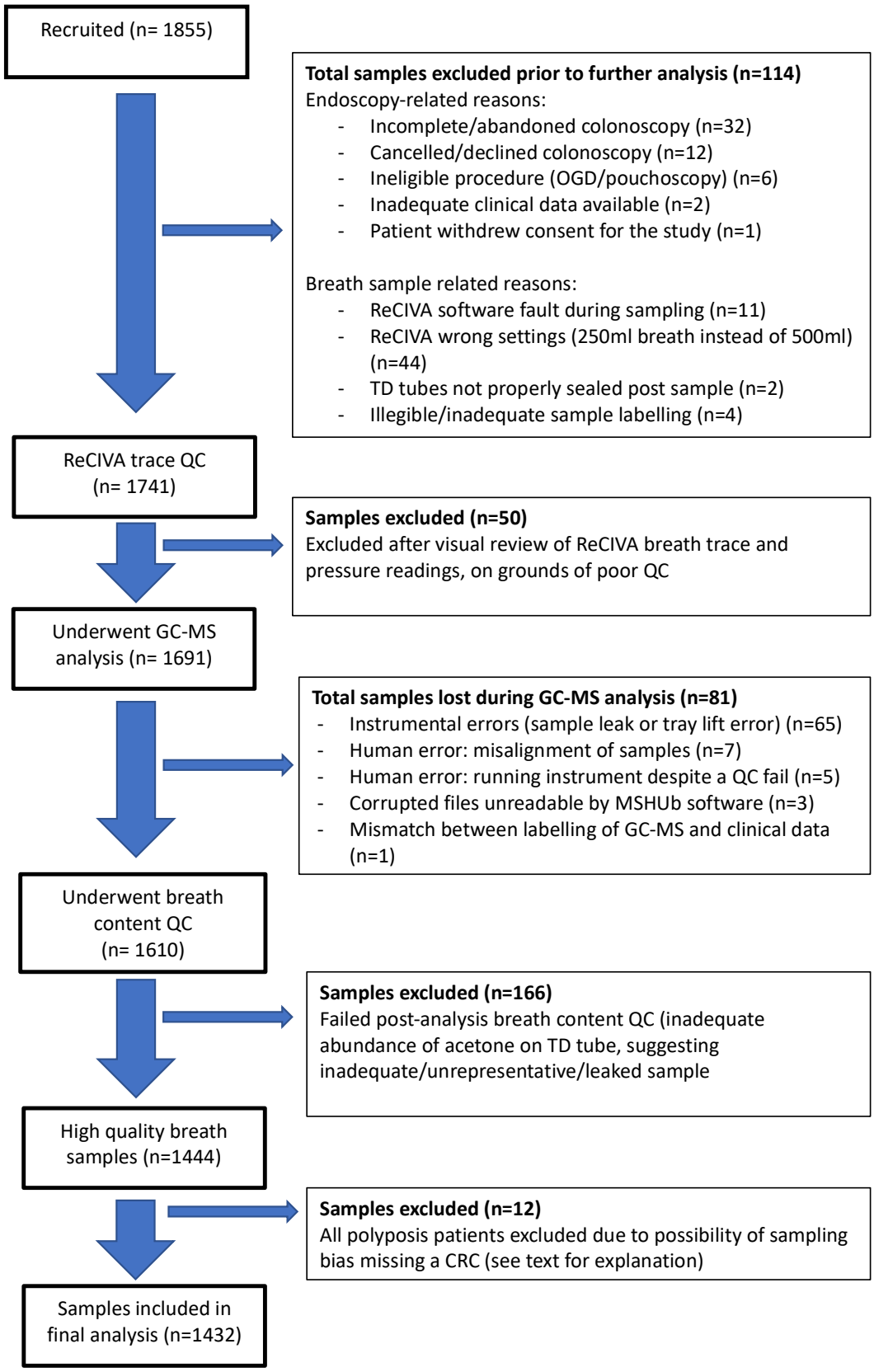


Figure 4-1 Flowchart outlining the reasons for post sampling data exclusion for the COBRA study, GC-MS dataset

4.4.2 Clinical details and demographics of recruited patients

Colonoscopy findings

1432 patients had breath samples analysed by GC-MS see Table 4-3 for their diagnoses and diagnostic group allocations. 162 patients were found to have CRC (11%), and 631 (43.7%) had colorectal polyps. As alluded to in the recruitment flowchart (see Figure 4-1), 12 polyposis patients were already excluded from this total on the basis that it was found to be too heterogeneous a group of patients; whilst some had >100 polyps present on the colonoscopy that day, others had only one 1 or 2 polyps due to very frequent surveillance and polypectomy. In addition, a significant number had had resections of a part of the colon already. Some were likely to have had upper gastrointestinal polyps also (unknown). Because of the variation in colonoscopic findings within this group, and the difficulty of confidently excluding a CRC in those with large numbers polyps, this group was excluded from the statistical analysis.

Table 4-3 Diagnosis at colonoscopy for 1432 patients

Group	Colonoscopic findings	Frequency	Percent of total
1	Normal	357	25%
2	Benign pathology	188	13%
3	IBD	106	7%
4	Low risk polyp(s)	348	24%
5	Intermediate risk polyp(s)	67	5%
6	High risk polyp(s)	204	14%
7	CRC	162	11%

Some patients were found to have more than one colonoscopic diagnosis and were grouped according to the most significant finding (diagnostic hierarchy is explained in the methods). However, in the case of IBD It is known that active IBD can alter the VOCs in breath (114) so this could represent a confounder. Patients coded into the cancer and polyp groups with additional concurrent IBD were therefore recorded. This was found to affect 13 patients only, 3 of which had inactive disease not visible on colonoscopy:

- Of 162 CRC patients, 5 had pre-existing IBD, where two had active UC visible on the colonoscopy, and 3 had inactive IBD not visible on colonoscopy.

- Of 204 high risk polyp grouped patients, only 2 had UC visible on the colonoscopy where one had been previously diagnosed.
- Of 67 intermediate risk polyp grouped patients, only 2 had previously-diagnosed UC visible on the colonoscopy.
- Of 348 low risk polyp grouped patients, 2 had previously-diagnosed UC visible on the colonoscopy and 2 had indeterminate colitis not previously diagnosed.

Demographics

See Table 4-4 for the breakdown of key characteristics of the 1432 patients included in the analysis. The control group (n=1270) included all positive and negative controls combined for the purposes of statistical comparison against the CRC group (n=162). 57.8% of the recruits to this study were male, with no significant difference in gender distribution between the CRC and control groups. CRC patients were significantly older than control patients, at 66.5 years in comparison to 63.0 years respectively ($p<0.001$). The majority of the patients were white British or European origin, and most were non-smokers, were current consumers of alcohol, with a median BMI of 26. There was no statistically significant difference between the distribution of these variables between the CRC and control groups. Although the median fasting time for CRC and control groups was similar, there was a statistically significant difference between groups, where the CRC group fasted for less time ($p<0.001$). The majority of patients had Moviprep as bowel preparation before their colonoscopy or theatre procedure. There was however a significant difference in bowel preparation distribution between cancer and control groups ($p<0.001$), largely because a narrower range of bowel preparations was used for pre-theatre patients and because 37 CRC patients had no bowel preparation at all before the breath test. The 'reason for colonoscopy/visit' and 'site sampled at' results were statistically significant between CRC and control groups because a significant proportion of CRC patients came from recruitment from theatres, which were located at St Mary's, West Middlesex and Chelsea and Westminster hospitals.

In the endoscopy unit, the study targeted BCSP patients primarily; 30 CRCs were detected at BCSP colonoscopies, giving a 4.5% CRC pick up rate from BCSP patients (30 out of 664 colonoscopies), lower than in the literature. (48) This was slightly less than the CRC pick-up rate in the 2WW patient group (5 out of 96 colonoscopies = 5.2%). Other CRCs were

detected in the surveillance (n=3), urgent symptoms (but not 2WW) (n=4) and re-scope for polyp removal groups (n=1), and none in the routine symptoms group. The rest of the cancer cases (n=119) were sampled pre-theatre having been identified for the study beforehand, representing an enriched cohort. As expected, the highest yield of polyp patients came from the BCSP and polyp surveillance groups; 65.3% of the polyp group were detected from the BCSP. The polyp pick-up rate was 61% in the BCSP patients (404 out of 664 BCSP colonoscopies), higher than in the literature. (48)

Past medical history and medication use were also recorded. There was a statistically significant difference in the number of patients who had had CRC in the past, making up 8.0% of the CRC group and 2.6% of the controls. These 13 patients therefore represented CRC luminal recurrence (in addition to extra-intestinal recurrence in some cases). Other statistically significantly increased co-morbid factors for the CRC group were the prevalence of known heart disease, laxative use, recent antibiotic use and warfarin (or other anticoagulant) use. Other comorbidities and medications used were comparable between CRC and control groups, see Table 4-4.

Table 4-4 Demographics of included patients, by main pathology groups

Clinical parameter		All patients, number, (%) (n=1432)	Cancer patients only number, (%) (n=162)	Polyp patients only number, (%) (n=619)	All Control (non-CRC) patients, number, (%) (n=1270)	P value between cancer and control patients
Gender	Female	599 (41.8%)	61 (37.7%)	216 (34.9%)	538 (42.4%)	p= 0.36
	Male	828 (57.8%)	101 (62.3%)	399 (64.5%)	727 (57.2%)	
	Unrecorded	5 (0.3%)	0	4 (0.6%)	5 (0.4%)	
Age (years)	Minimum	18	30	24	18	p< 0.001 **
	Maximum	90	90	87	87	
	Median (IQR)	64 (15)	66.5 (17)	66 (11)	63 (14)	
	Unrecorded age	7	0	2	7	
BMI (kg/m^2)	Minimum	14	18	15	14	p= 0.674
	Maximum	48	41	46	48	
	Median (IQR)	26 (8)	26 (7)	27 (7)	26 (8)	
	Unrecorded	697	36	318	661	
Ethnicity	Arab	45(3.1%)	5 (3.1%)	17 (2.7%)	40 (3.1%)	p= 0.568
	Asian / Asian British	220 (15%)	22 (13.6%)	85 (13.7%)	198 (15.6%)	
	Black / African / Caribbean / Black British	125 (8.7%)	19 (11.7%)	52 (8.4%)	106 (8.3%)	
	Other ethnicity	46 (3.2%)	8 (4.9%)	18 (2.9%)	38 (3.0%)	
	White British/ European	962 (67.2%)	104 (64.2%)	430 (69.4%)	858 (67.6%)	
	Unrecorded	34 (2.4%)	4 (2.5%)	17 (2.7%)	30 (2.4%)	
Smoking status	Current	185 (12.9%)	14 (8.6%)	101 (16.4%)	171 (13.5%)	p= 0.08
	Ex	442 (30.9%)	45 (27.8%)	199 (32.2%)	397 (31.3%)	
	Never	781 (54.6%)	102(63%)	311 (50.3%)	679 (53.5%)	
	Unrecorded	24 (1.7%)	1 (0.6%)	8 (1.3%)	23 (1.8%)	
Alcohol intake status	Current	890 (62.1%)	88 (54.4%)	418 (67.5%)	802 (63.1%)	p= 0.107
	Ex	73 (5.0%)	11 (6.8%)	35 (5.7%)	62 (4.9%)	
	Never	441 (30.7%)	61 (37.7%)	155 (25%)	380 (29.9%)	
	Unrecorded	28 (2.0%)	2 (1.2%)	11 (1.8%)	26 (2%)	
Bowel preparation taken	Citramag and Senna	111 (7.7%)	5 (3.1%)	55 (8.9%)	106 (8.3%)	p< 0.001**

	Moviprep	686 (47.9%)	106 (65.4%)	296 (47.8%)	580 (45.7%)	
	Klean prep	22 (1.5%)	0	10 (1.6%)	22 (1.7%)	
	Citrafleet	513 (35.8%)	13 (8%)	226 (36.5%)	500 (39.4%)	
	Picolax	55 (3.8%)	0	28 (4.5%)	55 (4.3%)	
	None	40 (2.8%)	37 (22.8%)	2 (0.3%)	3 (0.2%)	
	Unrecorded	5 (0.3%)	1 (0.6%)	2 (0.3%)	4 (0.3%)	
Time fasted (hours)	Minimum	4	4	9	7	p< 0.001**
	Maximum	72	36	52	72	
	Median (IQR)	24 (4)	22 (10)	24 (4)	24 (4)	
	Unrecorded	235	46	95	189	
Reason for colonoscopy/visit	BCSP screening	664 (46.4%)	30 (18.5%)	404 (65.3%)	634 (49.9%)	p< 0.001**
	Re-scope for polyp removal	27 (1.9%)	1 (0.6%)	20 (3.2%)	26 (2.0%)	
	Surveillance - IBD/polyps/FH	244 (17%)	3 (1.8%)	92 (14.8%)	241 (19.0%)	
	Symptoms - 2WW	96 (6.7%)	5 (3.1%)	28 (4.5%)	91 (7.2%)	
	Symptoms - urgent	127 (8.9%)	4 (2.5%)	33 (5.3%)	123 (9.7%)	
	Symptoms - routine	147 (10.3%)	0	38 (6.1%)	147 (11.6%)	
	Theatre resection patient for likely CRC	123 (8.6%)	119 (73.5%)	2 (0.3%)	4 (0.3%)	
	Unknown reason	4 (0%)	0 (0%)	2 (0.3%)	4 (0.3%)	
Site	Chelsea and Westminster	28 (2%)	26 (16.%)	1 (0.2%)	2 (0.2%)	p< 0.001**
	Charing Cross	366 (25.5%)	11 (6.8%)	184 (29.8%)	355 (28.0%)	
	Homerton	166 (11.6%)	8 (4.9%)	73 (11.8%)	158 (12.4%)	
	St George's	580 (40.5%)	14 (8.6%)	259 (41.8%)	566 (44.6%)	
	St Mark's	195 (13.6%)	8 (4.9%)	101 (16.3%)	187 (14.7%)	
	St Mary's	73 (5.1%)	71 (43.9%)	1 (0.2%)	2 (0.2%)	
	West Middlesex	24 (1.7%)	24 (14.8%)	0 (0%)	0 (0%)	
Past medical history	Previous CRC	46 (3.2%)	13 (8.0%)	10 (1.6%)	33 (2.6%)	p<0.001**
	Previous cancer excluding CRC	94 (6.5%) *	14 (8.6%)	41 (6.6%)	80 (6.3%)*	p= 0.257
Site of previous cancer(s)	Prostate	18 (1.3%)	3 (1.9%)	9 (1.5%)	15 (1.2%)	
	Skin	8 (0.6%)	1 (0.6%)	6 (1.0%)	7 (0.6%)	
	Bladder	5 (0.3%)	0	2 (0.3%)	5 (0.4%)	
	Head & neck	3 (0.2%)	1 (0.6%)	1 (0.2%)	2 (0.2%)	

	Hepatocellular	2 (0.1%)	0	2 (0.3%)	2 (0.2%)	
	Lung	5 (0.3%)	2 (1.2%)	1 (0.2%)	3 (0.2%)	
	Breast	14 (1.0%)	2 (1.2%)	5 (0.8%)	12 (0.9%)	
	Testicular	2 (0.1%)	0	2 (0.2%)	2 (0.2%)	
	Haematological	10 (0.7%)	0	3 (0.3%)	10 (0.8%)	
	Renal	2 (0.1%)	0	0	2 (0.2%)	
	Brain	1 (0.1%)	0	1 (0.2%)	1 (0.1%)	
	Other GI cancer	3 (0.2%)	0	1 (0.2%)	3 (0.2%)	
	Unknown type	35 (2.4%)	5 (3.1%)	8 (1.3%)	30 (2.4%)	
	Coeliac disease	3 (0.2%)	1 (0.6%)	0	2 (0.2%)	p= 0.228
	Past bowel resection	49 (3.4%)	5 (3.1%)	12 (1.9%)	44 (3.5%)	p= 0.803
	Barrett's oesophagus	6 (0.4%)	0	1 (0.2%)	6 (0.5%)	p= 0.381
	High blood pressure	491 (34.3%)	57 (35.2%)	261 (42.2%)	434 (34.2%)	p= 0.798
	Known heart disease	160 (11.2%)	28 (17.3%)	76 (12.3%)	132 (10.4%)	p= 0.009**
	Diabetes	217 (15.2%)	31 (19.1%)	97 (15.7%)	186 (14.6%)	p= 0.133
	Renal impairment	45 (3.1%)	8 (4.9%)	24 (3.9%)	37 (2.9%)	p= 0.164
	Chronic obstructive pulmonary disease	47 (3.3%)	6 (3.7%)	28 (4.5%)	41 (3.2%)	p= 0.749
	Asthma	103 (7.2%)	14 (8.6%)	44 (7.1%)	89 (7.0%)	p= 0.448
	Liver impairment	51 (3.6%)	3 (1.9%)	28 (4.5%)	48 (3.8%)	p= 0.212
Medications	Proton pump inhibitor	266 (18.6%)	31 (19.1%)	130 (21.0%)	235 (18.5%)	p= 0.846
	Non-steroidal anti-inflammatory	194 (13.5%)	18 (11.1%)	94 (15.2%)	176 (13.9%)	p= 0.336
	Laxative	38 (2.7%)	11 (6.8%)	17 (2.7%)	27 (2.1%)	p= 0.002**
	Antibiotics	26 (1.8%)	8 (4.9%)	10 (1.6%)	18 (1.4%)	p= 0.002**
	Ranitidine	18 (1.3%)	3 (1.9%)	8 (1.3%)	15 (1.2%)	p= 0.471
	Clopidogrel	29 (2.0%)	5 (3.1%)	18 (2.9%)	24 (1.9%)	p= 0.309
	Blood thinner (warfarin)	58 (4.1%)	14 (8.6%)	29 (4.7%)	44 (3.5%)	p= 0.002**
	Immunosuppressants	64 (4.5%)	10 (6.2%)	19 (3.1%)	54 (4.3%)	p= 0.265

*14 of these patients had >1 previous cancer
** significant p value at <0.05

4.4.3 Clinical details of colorectal cancer patients

Cancer specific details were recorded for all CRC patients. Most CRCs were left sided (62%), see Figure 4-2 and over half (64%) were late stage cancers T3 and T4, mostly with an N score of 0 to 1 and mostly without metastases, see Figure 4-3 to Figure 4-5. The range of size of tumour was 6mm to 130mm, median size 38.5mm (at greatest tumour diameter). 80% were moderately differentiated adenocarcinomas; see Figure 4-6.

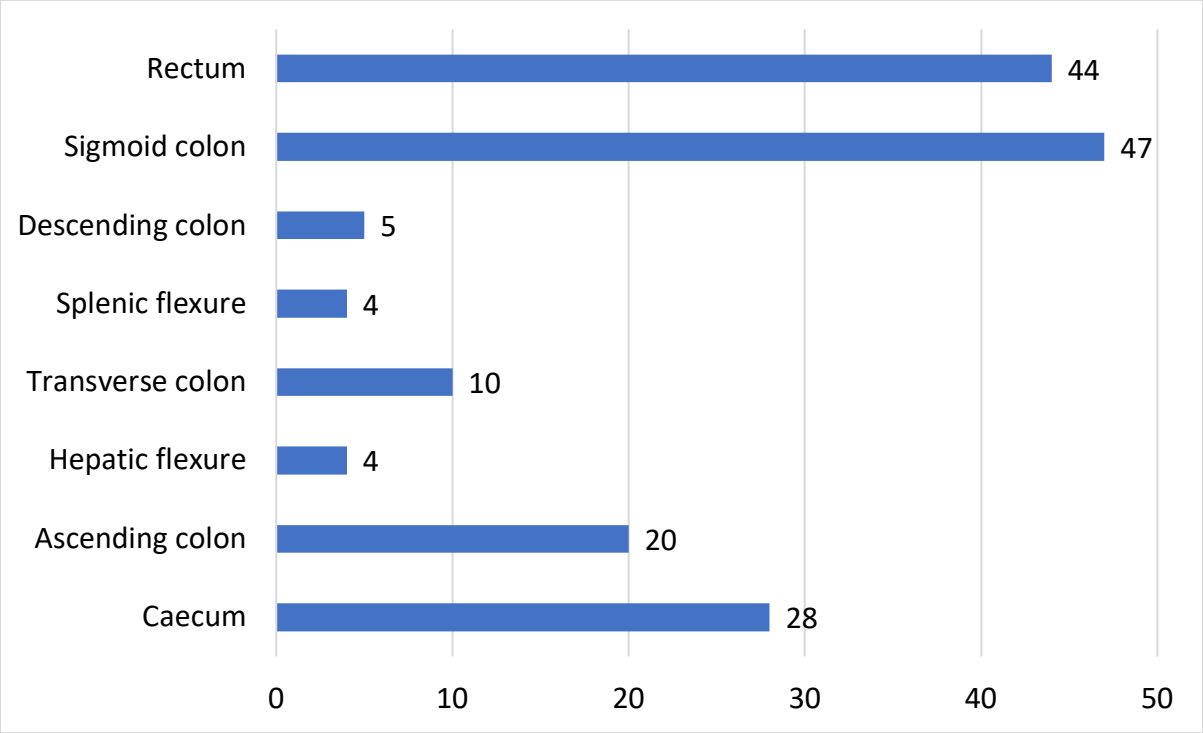


Figure 4-2 Frequency of site of 162 colorectal adenocarcinomas

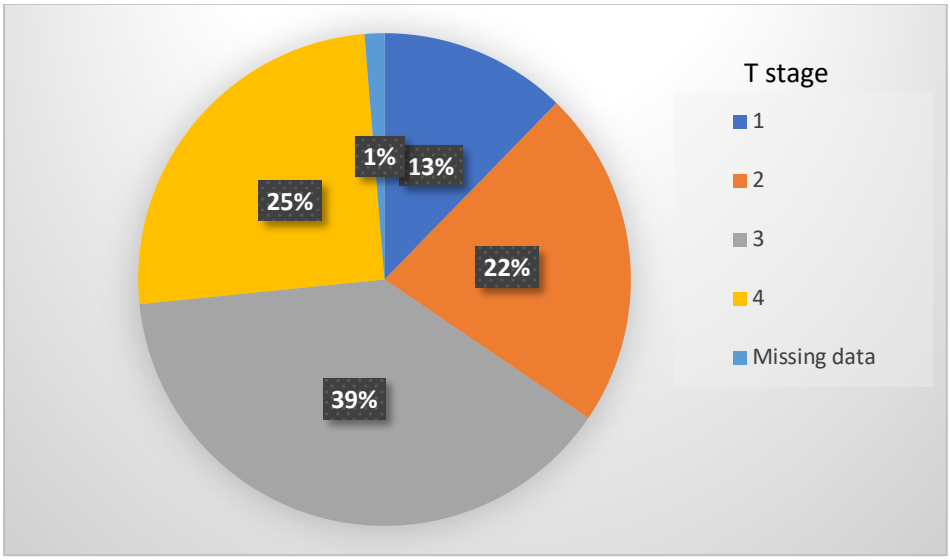


Figure 4-3 T stage score distribution in 162 colorectal adenocarcinomas

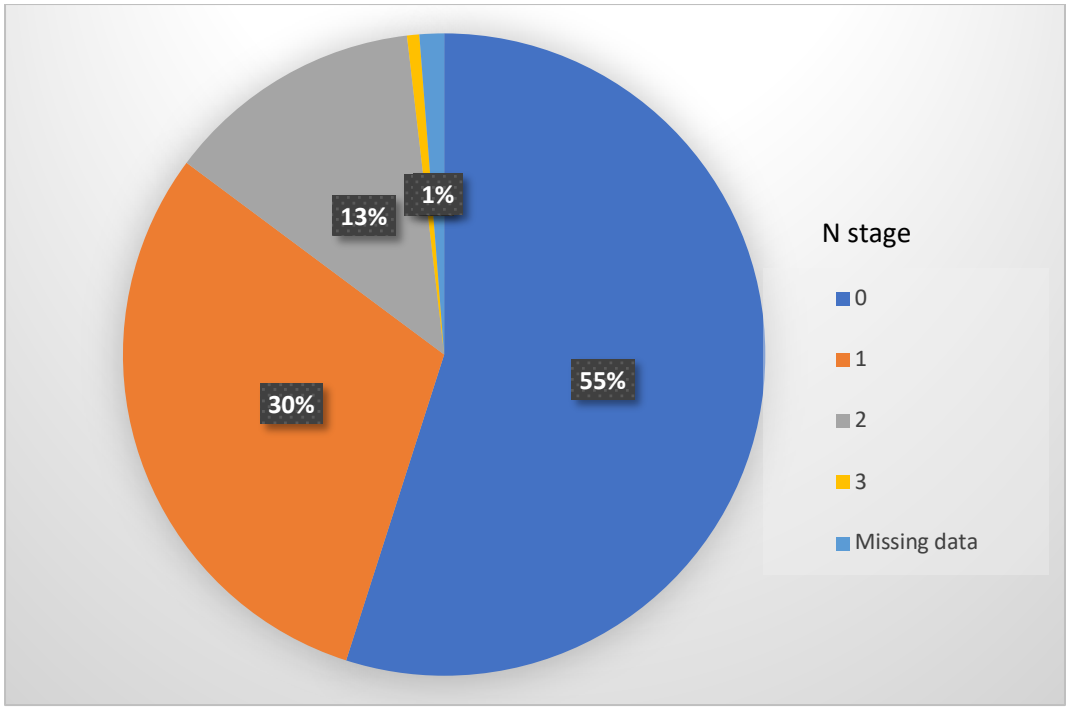


Figure 4-4 N stage score distribution in 162 colorectal adenocarcinomas

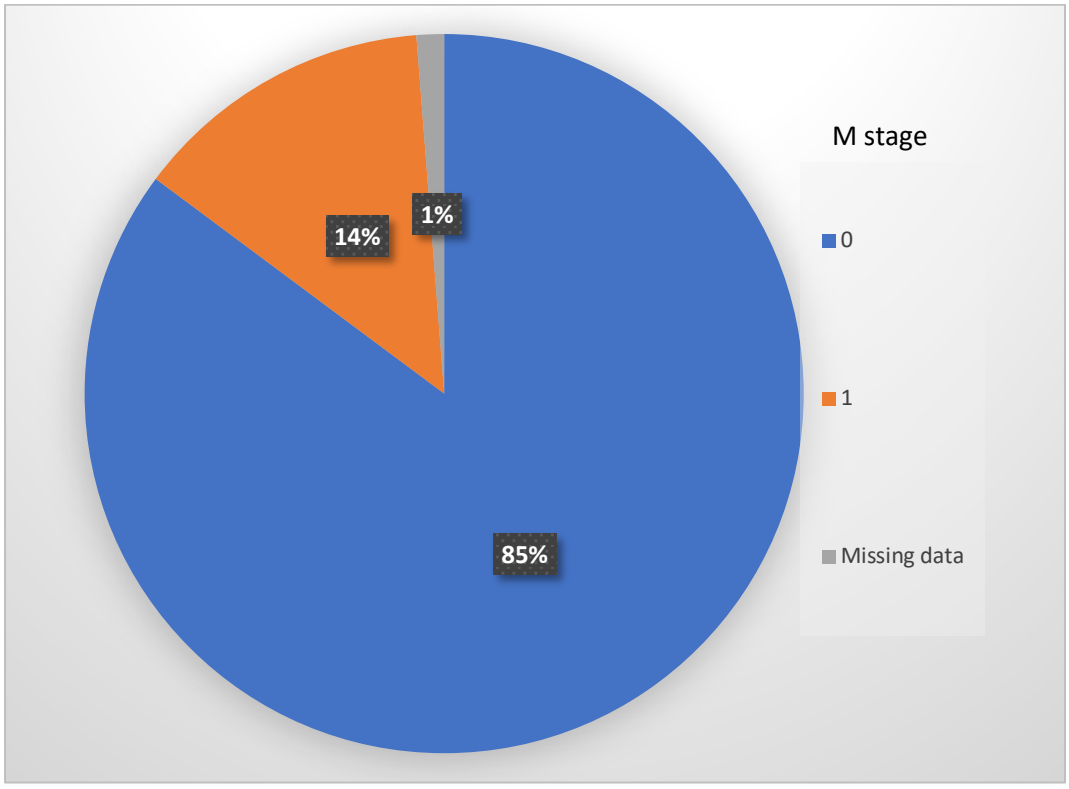


Figure 4-5 M stage score distribution in 162 colorectal adenocarcinomas

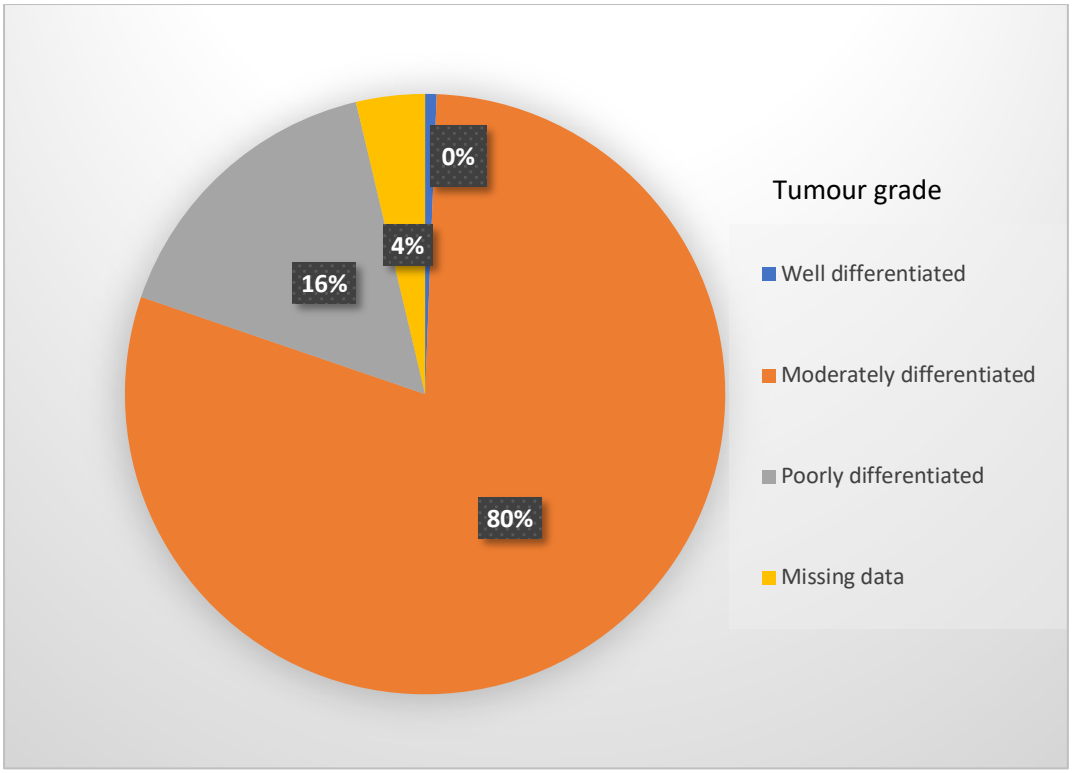


Figure 4-6 Frequency distribution of grade of tumour differentiation of 162 colorectal adenocarcinomas

The route of diagnosis of the CRC influenced what stage the CRC was. CRCs picked up in the BCSP were quite evenly distributed in terms of T stage, but the proportion of early cancers was higher in this group than any other (48% of BCSP cancers were T stage 1 or 2). This contrasted to the CRC patients who had been recruited via the theatre route; these patients tended to be symptomatic and a very high proportion of them (72%) had T stage 3 or 4 cancers, see Figure 4-7. This was an expected finding given that the BCSP is aimed at performing colonoscopy in asymptomatic individuals. In patients diagnosed with CRC, their age did not seem to necessarily correlate with the T stage that they were diagnosed at, see Figure 4-8.

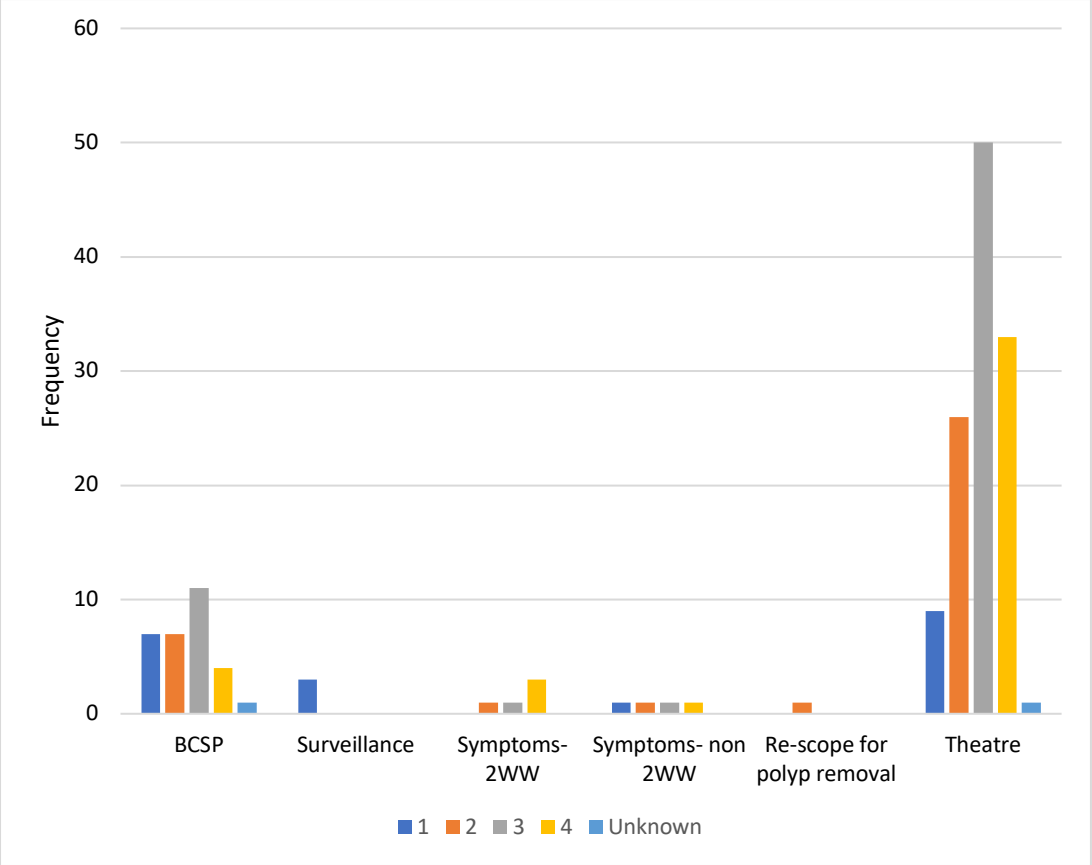


Figure 4-7 Distribution of T stage (1-4) at diagnosis of colorectal adenocarcinoma, by route of diagnosis

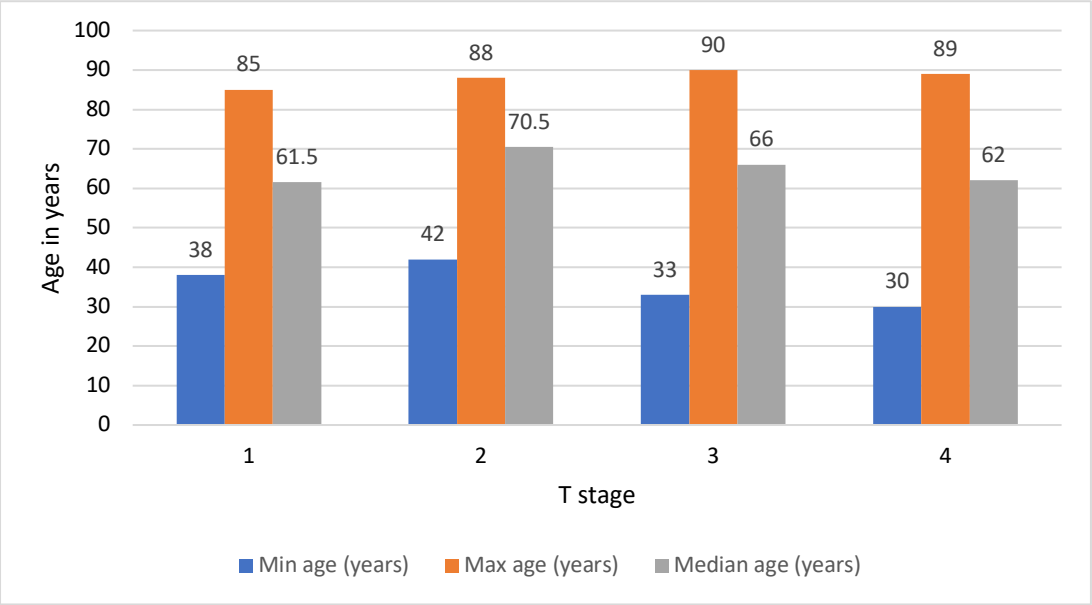


Figure 4-8 Age distribution by T stage, 162 colorectal adenocarcinomas

4.4.4 Sample processing times

The storage time of the cleaned TD tube prior to sampling, and the storage time of the breath sample on the TD tube before analysis by GC-MS are detailed in Table 4-5. There was no significant difference on a Kruskal-Wallis comparison between the 7 pathology groups with regards to storage of the TD tube prior to sampling (p=0.84), or post sampling (p=0.93), for all samples. No TD tubes were frozen prior to sampling, but post sampling 199 TD tubes were frozen for 1 to 114 days (median 17 days, standard deviation 40 days) before analysis, due to instrument down time/unavailability. For frozen tubes there was also no significant difference in post-sampling storage time between CRC and control groups (p=0.23).

Table 4-5 Thermal desorption tube storage time (days), n=1432

	Storage time (days)			
	Pre breath sampling	Post breath sampling		
	All samples (n=1432)	**All samples** (n=1432)	**Unfrozen samples only** (n=1233)	**Frozen samples only** (n=199)
Minimum time	0	0	0	1
Maximum time	20	114	8	114
Median time	3	1	1	17
Standard deviation	3	14	1	40

All tubes used for a patient sample (4 tubes per patient) were always conditioned/cleaned at the same time.

4.4.5 Results of initial univariate statistics

1024 features (VOCs) were identified in breath, and their peak area counts were tabulated by the MSHub programme. (171, 172) To start, a Kruskal-Wallis test was performed to determine if any of the 7 patient groups contained an abundance of a VOC that was statistically significant in differentiating between groups. 291 ions were found to be differentiating, with a $p<0.05$. A Mann Witney U analysis was performed for CRC (n=162) vs. control (n=1270) patients. 336 features (ions) were found to be differentiating, with a $p<0.05$. 95% of the features detected as discriminatory by the Kruskal-Wallis test were overlapping

with the features found by the Mann Witney U analysis, suggesting that it was the cancer group accounting for the significant differences in most cases. Groups were then interrogated in depth using advanced machine learning prediction models, where the clinical metadata was also incorporated as variables.

The figure below, Figure 4-9, shows the different machine learning analyses that were performed for the CRC analyses. The results of each of these analyses are explained subsequently.

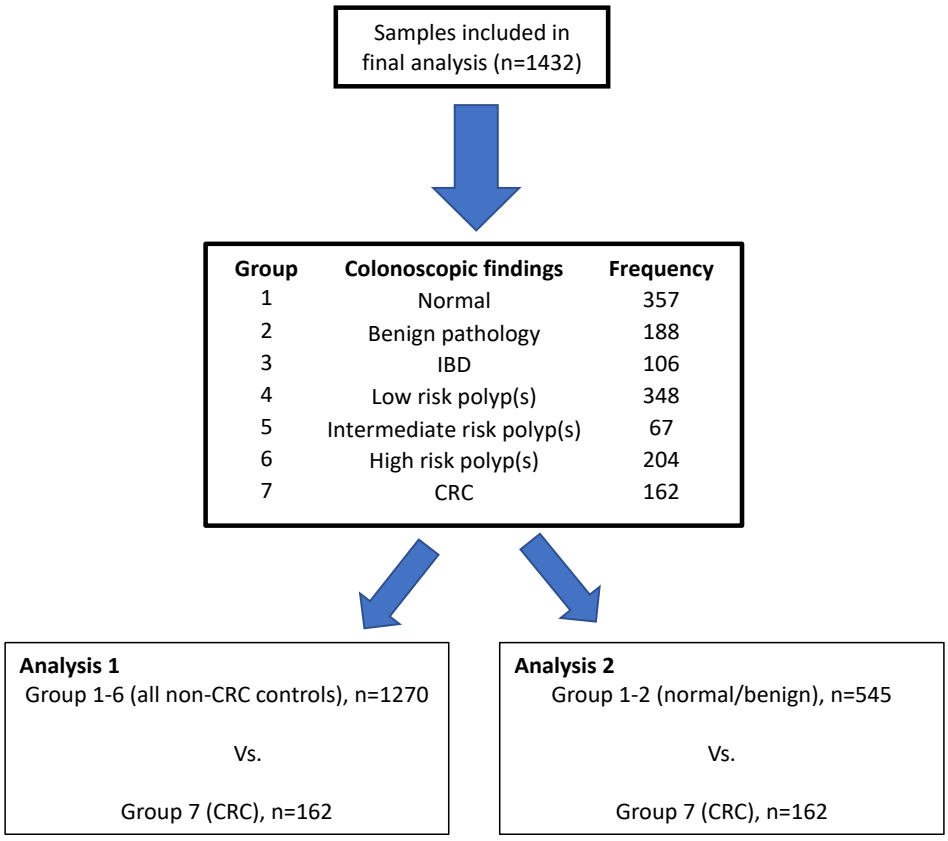


Figure 4-9 Machine learning analyses for CRC prediction

4.4.6 Results of machine learning prediction model, analysis 1: colorectal cancer vs. controls

The machine learning analysis was able to predict the CRC patients (n=162) from controls (n=1270) with a sensitivity of 0.77 (+-0.02), a specificity of 0.87 (+-0.01), a negative predictive value of 0.97 (+-0.00) and an accuracy of 0.86 (+-0.01), using machine learning elastic method. The area under the ROC curve was 0.87 (+-0.01); see Figure 4-10. The control (non-CRC) patient group comprised of both positive and negative controls; there were 357 negative controls (with normal colons at endoscopy), and 913 positive controls with benign

disease, IBD, or low/intermediate/high risk polyps at endoscopy. The ROC curve in Figure 4-10 was calculated based upon the results of a cross-validation method of 5 cycles of 5-fold stratified k-fold splits with reshuffling (explained in section 4.3.12). This means that the ROC curve was the average (mean) of ROC curves from individual runs, each with a slightly different feature selection and machine learning model. The area under the ROC curve for each individual cycle is shown in the key (Figure 4-10). Up to 99 features were used to generate this ROC curve, as determined by the machine learning algorithm, where "features" referred to individual ions but also individual clinical variables, i.e. any component that contributed to the separation of the groups. The number of features used was demonstrated by the RF selection score (the average number of times any given feature was selected for the method), where a score of 1 would mean that the feature in question was selected 100% of the time (also explained in section 4.3.12).

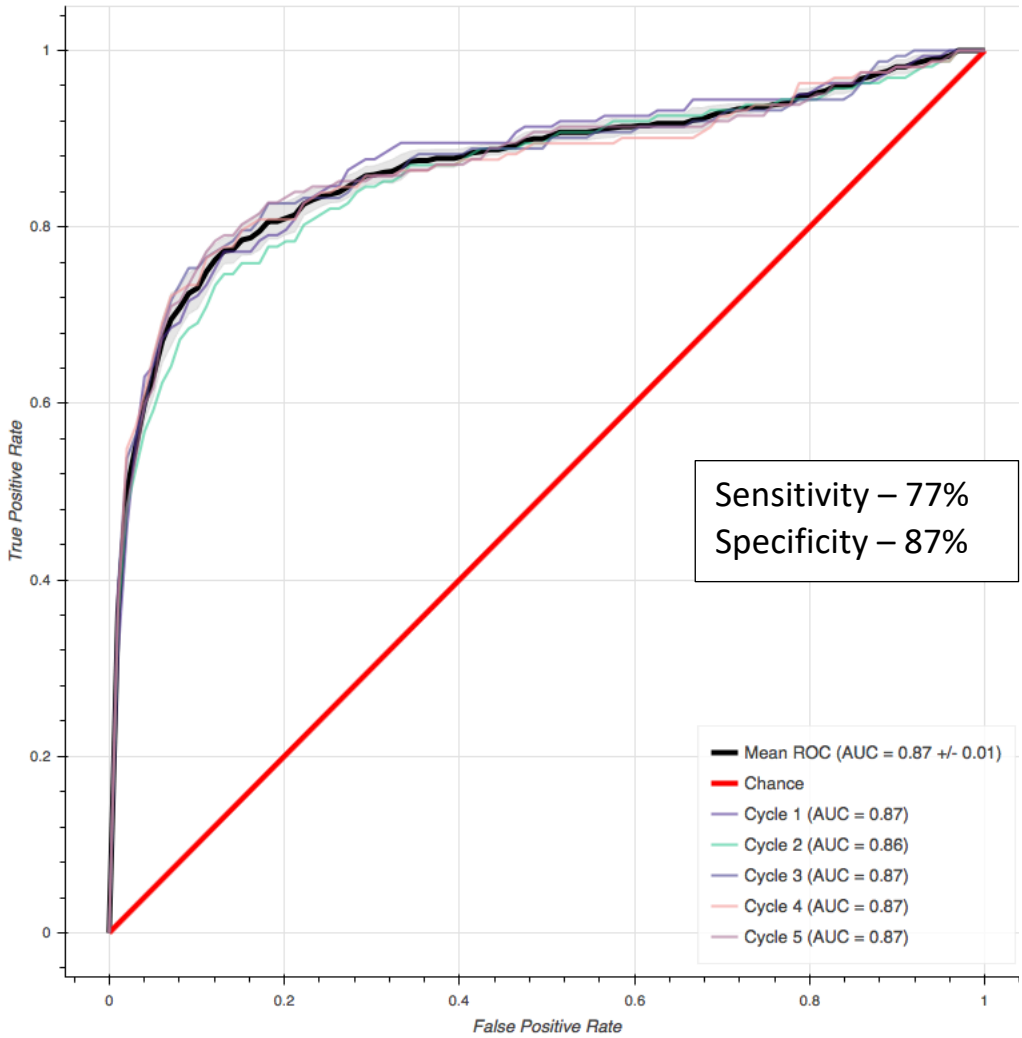


Figure 4-10 ROC curve for the prediction of CRC; CRC = 162, non-CRC = 1270

The top 25 chemical features, as well as 2 clinical features, that achieved the highest discriminatory scorings for CRC vs. non-CRC are listed below in Table 4-6. These were the features giving the highest contribution to the creation of the above presented ROC curve (Figure 4-10). Features were ranked using both RF selection and ANOVA, hence why the list of the top 25 ions was slightly different depending on which method was chosen. This was expected because ANOVA dealt with each feature one at a time and did not take feature cross-correlations or any other information into account, whereas RF constructed a model based on the entire ensemble of features and thus could take feature interactions into account. Both lists' features were interrogated. Features were identified by comparing the obtained mass spectra to possible matches suggested by the NIST database. (165) If the two mass spectra showed the same distribution and intensity of ions, then this was a match and the compound could be identified with a good degree of confidence. Where there was an imperfect but close spectral match, compound identification was tentative.

During the GC-MS analysis deconvolution, at any given retention time, a peak could be split into two (or more) peaks with different fragmentation patterns. The "percentage of peak" column in Table 4-6 shows how much of the original peak was explained by this new deconvolved peak. The lower the percentage, the less contribution this peak had and the less resolved it was. When the value was 100% there was a single peak, completely resolved. Any peaks that contributed to less than 20% of the original peak were excluded.

Table 4-6 List of top discriminating features contributing to the differentiation of CRC patients (n=162) from positive and negative control patients (n=1270)

Top discriminating features (VOCs) are ranked according to Random Forest (RF) and ANOVA feature selections (top 25 features from each method are listed).

Row	Rank using RF	Rank using ANOVA	Retention time (min)	% of peak	Tentative identification	Degree of spectral match	Comments	Possible biomarker?	Reference
1	1	7	8.09	34%	1,1,1,3,3,3-Hexafluoropropan-2-one, o-1,1,2,3,3,3-hexafluoropropyloxime	Moderate	Not found in any breath literature. All possible NIST matches were complex fluorinated compounds, so even if this is the incorrect exact identification, it is unlikely to be a biomarker.	No	
2	2	12	8.09	66%	1,1,2,2,3-pentafluoro-3-(1,1,2,2,2-pentafluoroethyl)cyclopropane	Not in NIST library	Not found in any breath literature. Fluorinated compound unlikely to be a biomarker. Cyclopropane is known to be an anaesthetic agent.	No	
3	3	1	6.54	100 %	Fluoromethyl pentafluoroprop-1-en-2-yl ether	Good	Not found in any breath literature. Fluorinated compound unlikely to be a biomarker.	No	
4	4	2	17.77	100 %	Unknown identity	Moderate	No good matches on NIST with regards to scores or spectra. The obtained mass spectrum had a base peak at m/z 99, and a peak at 149. Peak 149 is a well-known phthalate peak, where phthalates are found in plastics.	No	
5	5	6	22.02	100 %	Propyl ester of propanoic acid	Good	Has been found in breath literature. Used in perfumery and as a solvent, but has also has a role as a human metabolite deriving from propan-1-ol.	Yes	(177)
6	6	13	9.27	100 %	Dimethyl sulphide	Good	Has been found extensively in breath literature, largely as a biomarker of IBD. Has been shown to be raised in hepatocellular carcinoma and in gastric cancer tissue.	Yes	(160, 178)
7	7	49	39.02	77%	Beta isomethyl Ionone	Moderate	Flavour and fragrance agent, floral odour, found in perfumes	No	
8	8	3	24.65	27%	Methyl nonyl ester of carbonic acid	Moderate	Not found in any breath literature, high molecular weight (244) making in unlikely to be volatile. Also the ion only made up 27% of the peak area	No	
9	9	57	BMI					No	
10	10	58	40.12	88%	Alkane unknown	Good	NIST identifies this as hexadecane, but according to our alkane standard mix it is likely a branched C14 alkane, see discussion below. Alkanes can be endogenously produced, and are present in breath papers. Hexadecane has been shown to be a marker of oral squamous cell carcinoma and of ovarian cancer.	Yes	(157, 179)

11	11	23	42.48	100 %	Ethanol, 2-(2-butoxyethoxy)-, acetate	Good	Not found in any breath literature, high molecular weight making in unlikely to be volatile. Adhesive and binding agent.	No	
12	12	8	6.68	24%	1,3-Dioxolane-2-methanol	Good	Found to be higher in the breath of gastric cancer patients using SPME GCMS	Yes	(180)
13	13	24	10.89	72%	2-Propenenitrile	Good	Used to make other chemicals such as plastics, synthetic rubber, and acrylic fibres.	No	
14	14	204	40.52	86%	2-Phenoxy-ethanol	Good	Used as a fixative for perfumes and as a bactericide. Found in 100% of breath samples in a paper comparing smokers and non-smokers (n=119)	Yes	(181, 182)
15	15	105	4.75	100 %	Cyclopropane	Good	Cyclopropane has been found in the breath of healthy patients, and higher breath levels have been associated with breast cancer.	Yes	(95, 183)
16	16	5	31.14	27%	3,4-dimethyl-1,5-Cyclooctadiene,	Good	Likely to be exogenous, although similar compounds were possibly found in metabolites of a plant leaf.	Yes	(184)
17	17	9	23.53	31%	2-propenyl ester of acetic acid	Good	Used in cosmetics and toiletries. Could be related to allyl alcohol, which is found in garlic and in breath.	Yes	(185)
18	18	719	39.02	23%	1-Propylnonylbenzene	Good	Not found in any breath literature, high molecular weight making in unlikely to be volatile - could possibly be an artefact from the carbograph/tenax sorbent. Also, the ion only made up 23% of the peak area.	No	
19	19	10	32.69	39%	Similar to 2-propenyl ester of acetic acid		Overlapping ester to 2-propenyl ester of acetic acid.	Yes	
20	20	19	35.12	100 %	(chlorocyanofluoromethyl)phenyl-Mercury	Moderate	Not found in any breath literature. All NIST matches were complex fluorinated compounds, even if this is incorrect ID it is unlikely to be a biomarker	No	
21	21	15	7	69%	Chloroethane	Good	Used in paints, cosmetics, aerosols, anaesthetics, not known to occur as a natural product.	No	
22	22	11	20.41	32%	3-methyl-butanenitrile	Good	Present in tomato and tobacco plants. Significantly higher in smokers in this dataset, so can't be a reliable biomarker, see below.	No	(186)
23	23	69	31.71	57%	2,2,4-Trimethyl-3-pentanol	Good	Present in spices and flavourings, as well as alcoholic beverages and fruits	Yes	(187, 188)
24	24	257	38.74	60%	Alkane unknown	Good	NIST identifies this as hexadecane, but according to our alkane standard mix it is likely a branched C14 alkane, see discussion below. Alkanes can be endogenously produced, and are present in breath papers. Hexadecane has been shown to be a marker of oral squamous cell carcinoma and of ovarian cancer.	Yes	(157, 179)
25	25	145	Age					No	
26	41	4	17.72	72%	2-Propanol, 1,1,1,3,3,3-hexafluoro-	Good	AKA sevoflurane: sweet-smelling, non-flammable, highly fluorinated methyl isopropyl ether used for induction and maintenance of general anaesthesia.	No	(188)
27	105	14	13.64	21%	Ethyl methylphosphinate	Moderate	Component in various medications- osteoporosis, opioids, drugs for central nervous system disorders, anti-epileptics.	No	(189)
28	95	16	22.24	41%	Methyl 2-butynoate	Moderate	Is a component of ((1S,4S,5R,6R,9S,10R,12R,14R)-4,6-dihydroxy-7-(hydroxymethyl)-3,11,11,14-tetramethyl-15-oxotetracyclo[75,101,5010,12]pentadeca-2,7-dien-5-yl (2 Z)-2-methyl-2-butenoate), 'milk-weed', thought to have anti-cancer properties.	Yes	(190)

29	54	17	32.41	76%	Diethylene Glycol	Good	Used in the manufacture of polyester resins, polyurethanes, and plasticisers.	No	
30	65	18	36.92	27%	2-Pyrrolidone	Good	Used in inkjet cartridges.	No	(191)
31	56	20	24.65	66%	1-Undecanol	Good	Endogenous and comes from ingestion of fruits and plants. Known to be excreted in urine.	Yes	(188)
32	62	21	3.81	98%	Unknown identity	NA	Likely artefact as eluting very early (RT of 3.81)	No	
33	125	22	18.11	90%	Phenol	Good	Originates from foods (coffee and cocoa), but is also used commonly used as an antiseptic and disinfectant.	Yes	(188)
34	83	25	21.26	29%	3-ethyl-hexane	Good	Found as a biomarker for lung cancer	Yes	(192)

From the comprised list (Table 4-6) of 25 top CRC-differentiating ions from two different machine learning prediction models (ANOVA and RF), a short list was created. These short-listed ions were manually selected with the following criteria: (i) the VOC could be considered endogenous; (ii) the VOC had a physiological role that could explain its involvement in CRC. 3-methyl-butanenitrile was not considered as a compound of potential importance because of its presence in tobacco plants, (186) leading to interrogation of the COBRA dataset; there was a significantly higher abundance of this compound in smokers (n=185) compared to non-smokers (n=781), p=000009 using a Mann Whitney U test. The identification of this compound using the NIST library showed a good degree of confidence, since we obtained a good spectral overlap. 3-methyl-butanenitrile was therefore excluded as a potential CRC marker. Table 4-7 details the 15 ions that were taken forward as potential VOC biomarkers for further investigation, with their statistical scorings.

Table 4-7 Top 15 cancer-differentiating ions, defined as those with the potential to be CRC biomarkers

VOC number (in order of predictive importance)	Retention time (mins)	% of peak	RF score	RF Selection score	p value (from ANOVA analysis)	Tentative identification	CAS number	Known CRC biomarker?
1	22.02	100%	0.01	1	0.000	Propyl ester of propanoic acid (propyl propionate)	106-36-5	No
2	9.27	100%	0.01	1	0.000	Dimethyl sulphide	75-18-3	No
3	40.12	88%	0.005	0.84	0.000	Alkane	Unknown	Yes- in breath (84, 116, 193)
4	6.68	24%	0.004	0.76	0.000	1,3-Dioxolane-2-methanol	5694-68-8	No
5	40.52	86%	0.004	0.68	Not in ANOVA top 100	2-Phenoxy-ethanol	122-99-6	No
6	4.75	100%	0.004	0.760	Not in ANOVA top 100	Cyclopropane	75-19-4	No
7	31.14	27%	0.004	0.72	0.000	3,4-dimethyl- 1,5-Cyclooctadiene	21284-05-9	No
8	23.53	31%	0.004	0.72	0.000	2-Propenyl ester of acetic acid (allyl acetate)	591-87-7	No
9	32.69	39%	0.003	0.6	0.000	Overlapping ester similar to allyl acetate	591-87-7	No
10	31.71	57%	0.003	0.56	0.000	2,2,4-Trimethyl-3-pentanol	5162-48-1	No
11	38.74	60%	0.002	0.44	Not in ANOVA top 100	Alkane	Unknown	Yes- in breath (84, 116, 193)
12	22.24	64%	0.000	0.08	0.000	Methyl 2-butynoate	23326-27-4	No
13	24.65	66%	0.001	0.2	0.000	1-Undecanol	112-42-5	No
14	18.11	90%	Not in RF top 100	0.04	0.000	Phenol	108-95-2	No
15	21.26	29%	0.000	0.08	0.000	3-Ethyl-hexane	619-99-8	Yes- in plasma (126)

RF denotes random forest. RF score and RF Selection score are explained in section 4.3.12.

As explained above in section 4.3.12, the RF features scorings usually should add to 1, when all features that contributed to the model are included. However, Table 4-7 demonstrates the RF score for the top 15 most promising and realistic biomarker candidates only, hence why the RF scores depicted here do not sum to 1. The numerical values of the scores can only be interpreted in relation to the other feature scores, rather than as absolute values. True important features should have high scores relative to the noise features. VOC 12 and 15 (methyl 2-butynoate and 3-ethyl-hexane) both have a RF score of 0, however they are included still because they were within the top 25 features found in the ANOVA analysis (even if not featuring highly in the RF analysis).

With regards to the selection scores of the top 15 features shown in Table 4-7, scores ranged from 4% to 100%. This may represent a limitation of the data, specifically for VOC 12, 14 and 15, (methyl 2-butynoate, phenol and 3-ethyl-hexane), representing 8%, 4% and 8% selection scores respectively.

Applying just these top 15 features in isolation on the dataset, an area under the ROC curve of 0.83, and a 95% confidence interval of 0.79-0.86 was obtained, see Figure 4-11.

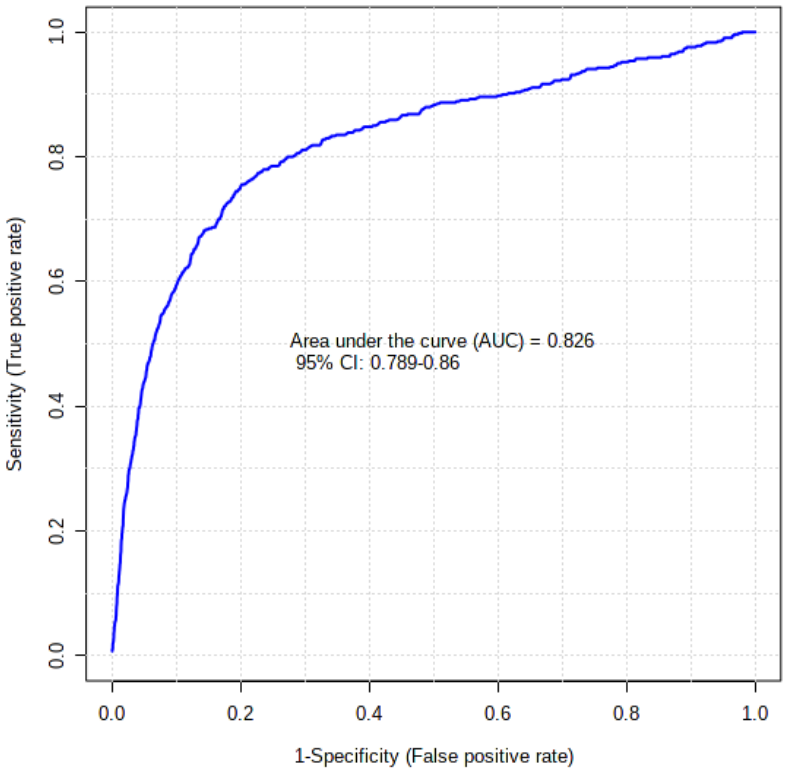


Figure 4-11 ROC curve illustrating predictive power of 15 VOCs in determining CRC from non-CRC

4.4.7 Results of machine learning prediction model, analysis 2: colorectal cancer vs. no colorectal pathology

The same machine learning analysis as performed above was repeated for the CRC group (n=162) vs. the normal/benign colorectal pathology group only (n=545) (i.e. excluding IBD and polyps). This group had either normal colonoscopies or benign findings such as a haemorrhoid, diverticular disease or a benign non-IBD associated anal fissure. Interestingly, 23 of the resultant top 25 features using RF selection overlapped with the top 25 features for the larger CRC vs. non-CRC comparison described above, suggesting that the markers found could be truly CRC-specific and unaffected by other colorectal pathologies such as IBD and polyps. The two new VOCs that were not in the pre-existing list were pentafluoroethane, similar to other fluorinated compounds found in the CRC vs. non-CRC comparison, see Table 4-6, and 2-methyl-2-propanol, an alcohol.

Each of the top 15 discriminating ions for CRC was explored in detail, within its appropriate chemical group.

4.4.8 The esters

VOCs 1, 8, 9 and 12 were tentatively identified as propyl propionate, allyl acetate, a similar overlapping ester to allyl acetate, and methyl 2-butynoate. See appendix 10 for the mass spectrum for each VOC and the comparative spectrum from the NIST database, to show how these compounds were identified.

All four of the obtained esters were present in significantly higher abundance in the breath of patients with CRC (n=162) compared to those without CRC (n= 1270). The ion peak area counts are given for both groups in Table 4-8, and representative boxplots of the distributions in each group are demonstrated in Figure 4-12 to Figure 4-15. For each boxplot, the median is represented by the solid horizontal line, the whiskers represent the minimal and maximal value, and the box represents the interquartile range.

Table 4-8 Abundance, measured in peak area count, of the four significant esters measured by GC-MS between patients with (n= 162) and without CRC (n=1270)

	VOC1: Propyl propionate		VOC 8: Allyl acetate		VOC9: Overlapping ester similar to allyl acetate		VOC 12: Methyl 2-butynoate	
	Non-CRC	**CRC**	**Non-CRC**	**CRC**	**Non-CRC**	**CRC**	**Non-CRC**	**CRC**
Median abundance	850.16	3983.54	12.98	3239.60	1098.42	29337.06	115.62	211.43
Min. abundance	0.00	16.68	0.00	0.00	0.00	0.00	0.00	0.00
Max. abundance	1122250.52	2388798.58	279440.10	249409.0	222161.0	147909.50	2344087.89	2186696.61
Standard deviation	60937.50	212372.46	27445.07	37679.17	26744.41	33364.88	295073.60	549823.90
Interquartile range	3990.03	44371.15	1261.99	45853.17	7838.0	53937.68	216.61	95368.85
Number of samples with zero peak area (% of total)	103 (7.2%)	9 (0.6%)	534 (37.2%)	27 (2.0%)	46 (3.2%)	2 (0.2%)	39 (2.7%)	4 (0.3%)

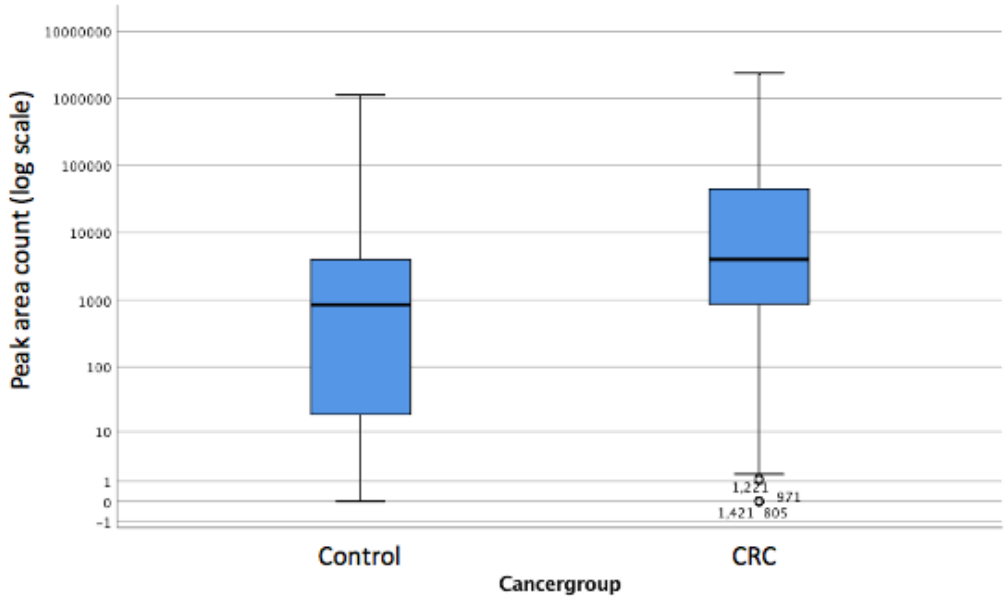


Figure 4-12 Boxplot of propyl propionate abundance, VOC 1, in breath in non-CRC vs. CRC.

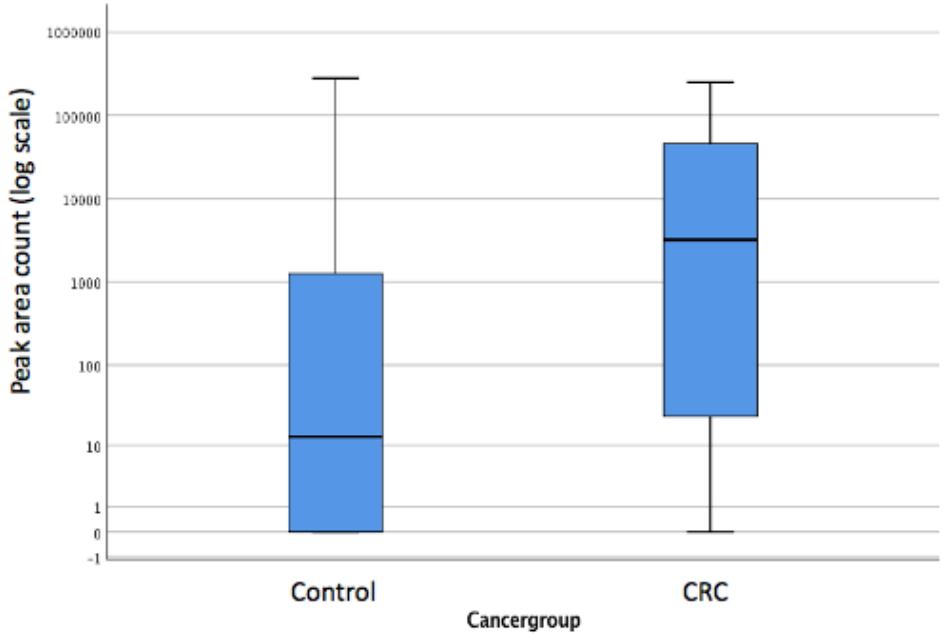


Figure 4-13 Boxplot of allyl acetate abundance, VOC 8, in breath in non-CRC vs. CRC.

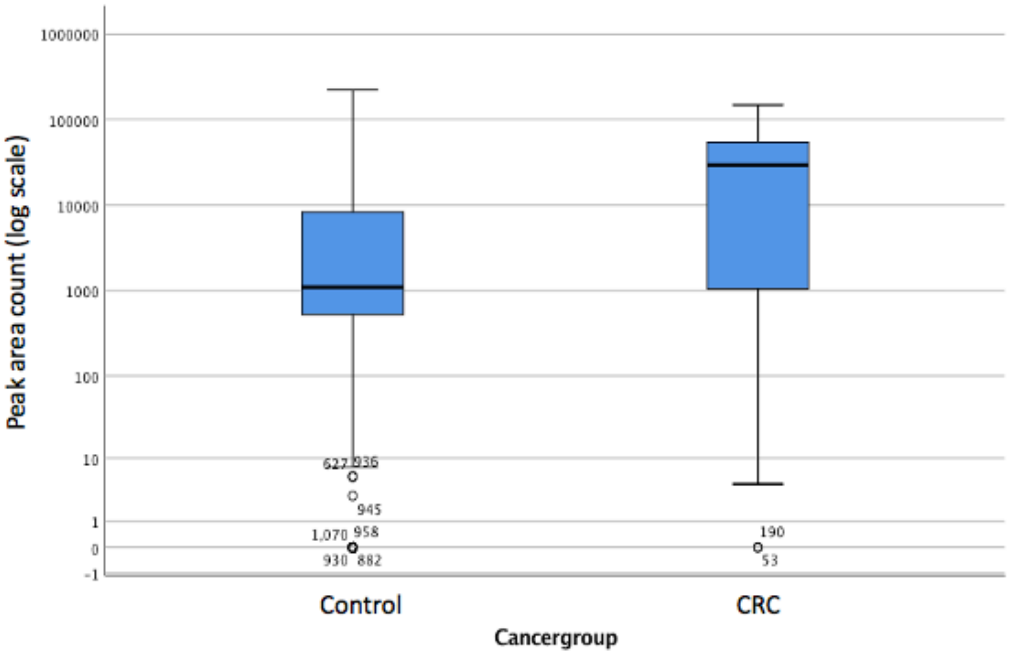


Figure 4-14 Boxplot of overlapping ester to allyl acetate abundance, VOC 9, in breath in non-CRC vs. CRC.

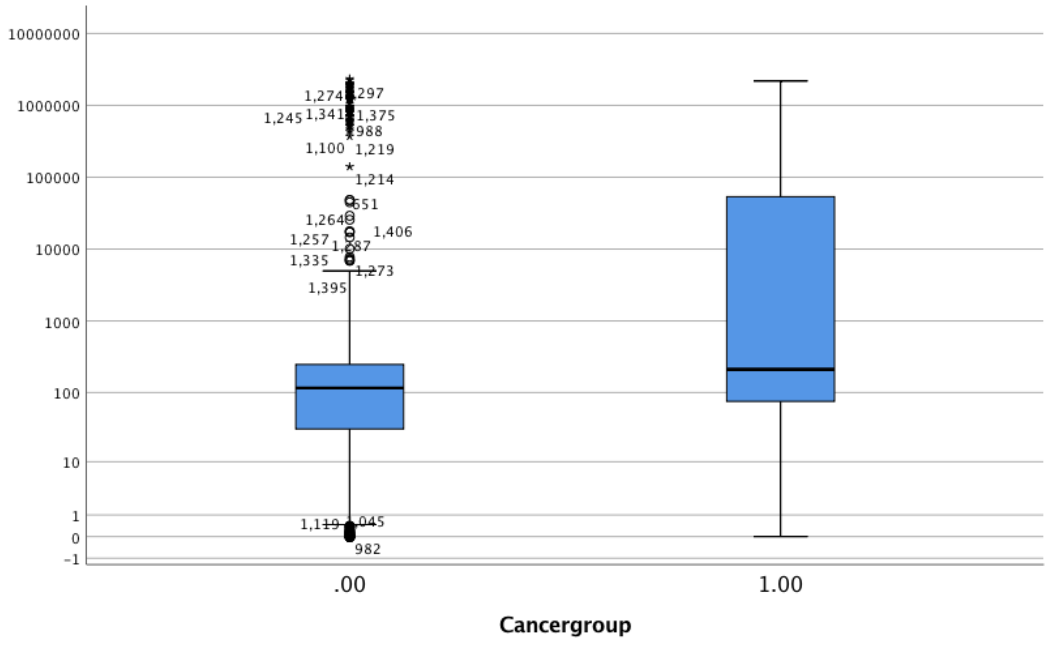


Figure 4-15 Boxplot of methyl 2-butynoate abundance, VOC 12, in breath in non-CRC vs. CRC.

4.4.9 Sulphur compounds

VOC 2 was identified as dimethyl sulphide, with a good match between the obtained mass spectrum and the NIST database, see appendix 11. It has a chemical formula of $C_2H_6S_2$ and an m/z of 63. Dimethyl sulphide was found to be present in a significantly higher abundance in the breath of patients with CRC (n=162) compared to those without CRC (n= 1270). The obtained peak area count for the two study groups are given for both groups in Table 4-9, and representative boxplots of the distributions in each group are demonstrated in Figure 4-16. The median is represented by the solid horizontal line, the whiskers represent the minimal and maximal value, and the box represents the interquartile range. The boxplot shows that the abundance of dimethyl sulphide was higher in CRC patients, but that there was some overlap.

Table 4-9 Abundance, measured in peak area count, for dimethyl sulphide measured by GC-MS between patients with (n= 162) and without CRC (n=1270)

	Dimethyl sulphide	
	Non-CRC group	**CRC group**
Median abundance	53389.58	74448.61
Minimum abundance	0.00	16.68
Maximum abundance	2066857.40	1974132.90
Standard deviation	123533.43	310168.72
Interquartile range	82034.83	168510.46
Number of samples with zero peak area (% of total samples)	17 (0.01%)	0 (0%)

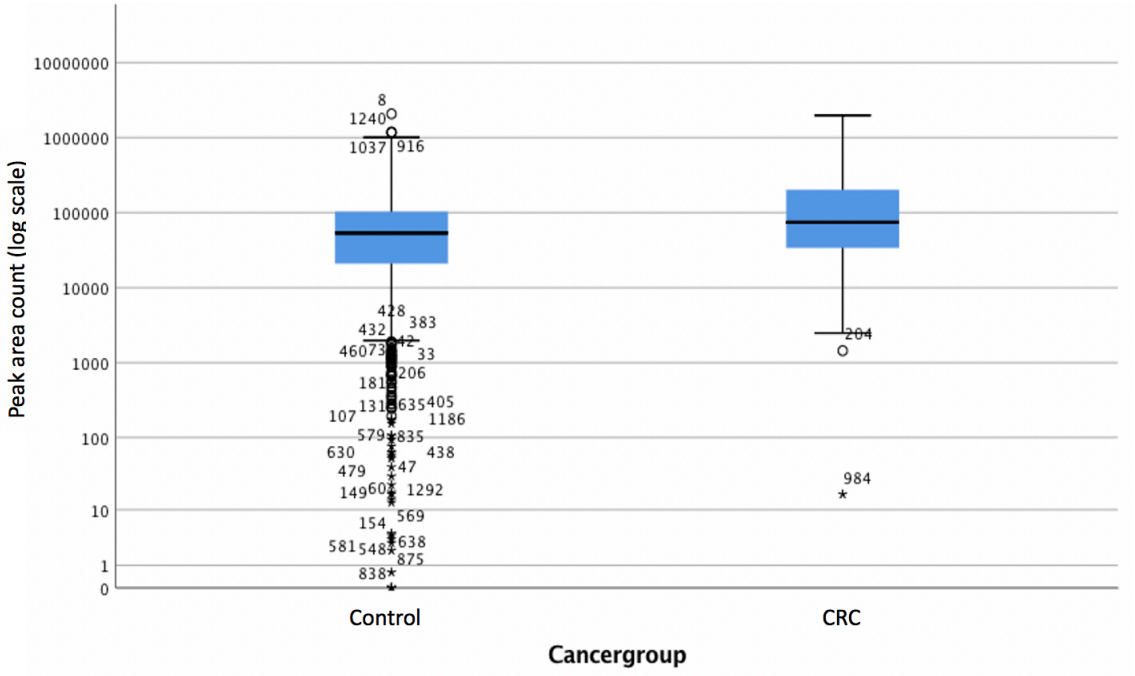


Figure 4-16 Boxplot of dimethyl sulphide abundance, VOC 2, in breath in non- CRC vs. CRC.

4.4.10 The alkanes

VOCs 3, 11 and 15 were identified as two unidentified alkanes and 3-ethyl-hexane respectively. All three of these alkanes were significantly lower in CRC patients than in non-CRC patients. Appendix 12 shows the mass spectrum for each VOC and the comparative spectrum from the NIST database, to show how these compounds were identified. However, alkanes are notoriously difficult to identify as the mass spectra are very similar using GC-MS (as demonstrated by the mass spectra of VOC 3 and 11), so the spectra alone are not enough to be able to give unequivocal identification. To aid with this, a standard mix of 12 straight chain alkanes, from C8 to C20 (octane, nonane, decane etc.) was analysed by GC-MS to obtain specific retention times, for identification purposes. Retention time is dependent upon volatility and affinity for the column, where more volatile compounds will have a lower retention time. The retention times for the alkane standards were, as expected, aligning in sequence as molecules became less volatile. The retention time peaks for the two unidentified alkanes discovered in the COBRA study fell between the retention time peaks for C13 and C14 alkanes. This makes them very likely to be C14 alkanes, but with a branched carbon chain, causing them to elute from the column slightly earlier than the C14 unbranched

alkane, as they are slightly less retentive due to their stereochemistry. The conclusion was therefore that both VOCs are likely to be branched chain alkanes of C14.

All three alkanes were found to be present in significantly lower abundance in the breath of patients with CRC (n=162) compared to those without CRC (n= 1270). The obtained peak area count for the two study groups are shown in Table 4-10 and representative boxplots of the distributions in each group are demonstrated in Figure 4-17 to Figure 4-19.

Table 4-10 Abundance, measured in peak area count, of three significant alkanes measured by GC-MS between patients with (n= 162) and without CRC (n=1270)

	VOC 3: C14 branched alkane		**VOC 11: C14 branched alkane**		**VOC 15: 3-ethyl-hexane**	
	Non-CRC	**CRC**	**Non-CRC**	**CRC**	**Non-CRC**	**CRC**
Median abundance	98217.08	43731.18	204236.0	115050.40	69474.55	10455.48
Minimum abundance	0.00	0.00	0.00	0.00	0.00	0.00
Maximum abundance	1154993.47	479693.97	2605583.20	965395.20	1629179.98	396575.10
Standard deviation	130512.14	108128.11	239753.50	216157.40	142507.00	84657.72
Interquartile range	114171.01	124357.77	204673.30	227899.30	140029.80	62747.03
Number of samples with zero peak area (% of total samples)	10 (0%)	3 (0%)	10 (0%)	3 (0%)	170 (11.9%)	58 (4.1%)

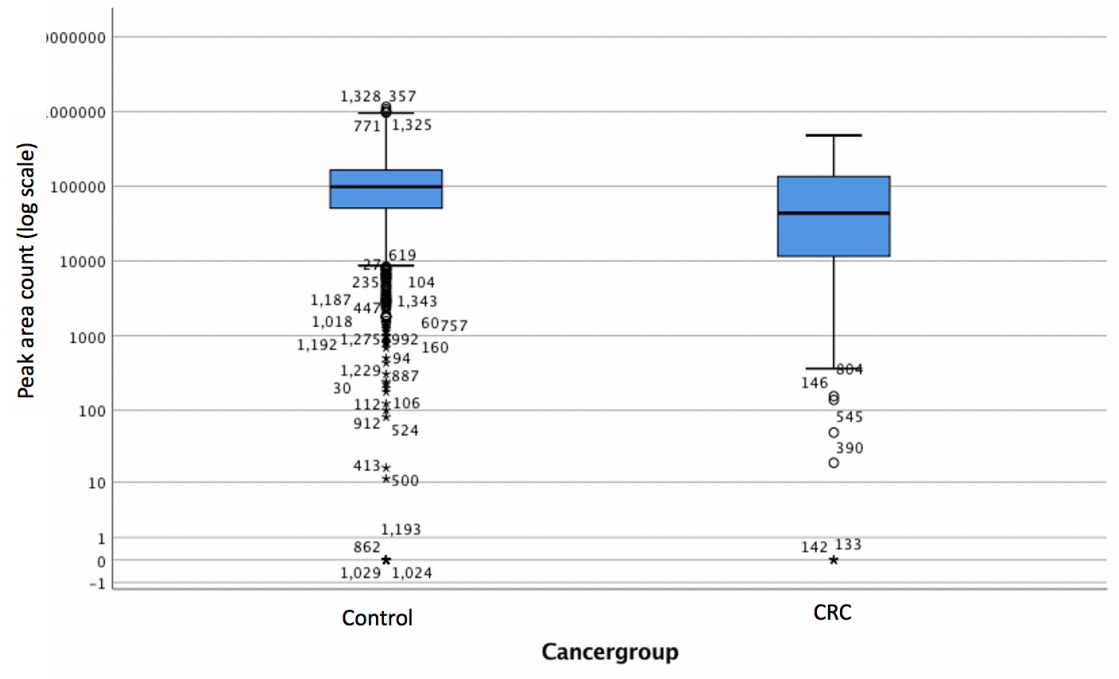


Figure 4-17 Boxplot of alkane biomarker abundance, VOC 3, in breath in non-CRC vs. CRC.

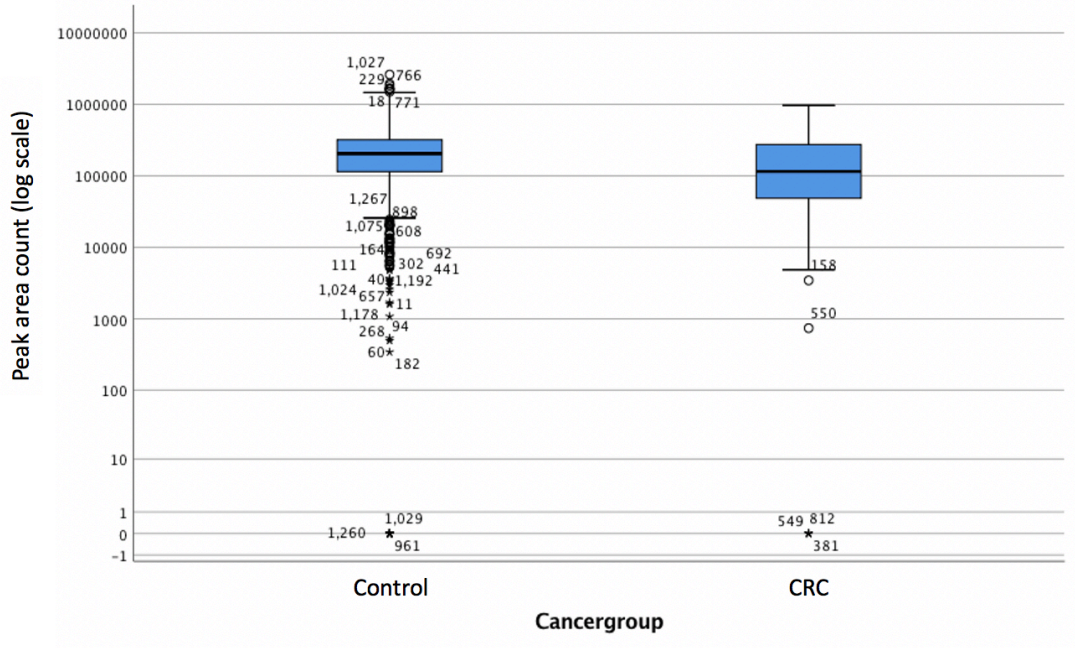


Figure 4-18 Boxplot of alkane biomarker abundance, VOC 11, in breath in non-CRC vs. CRC.

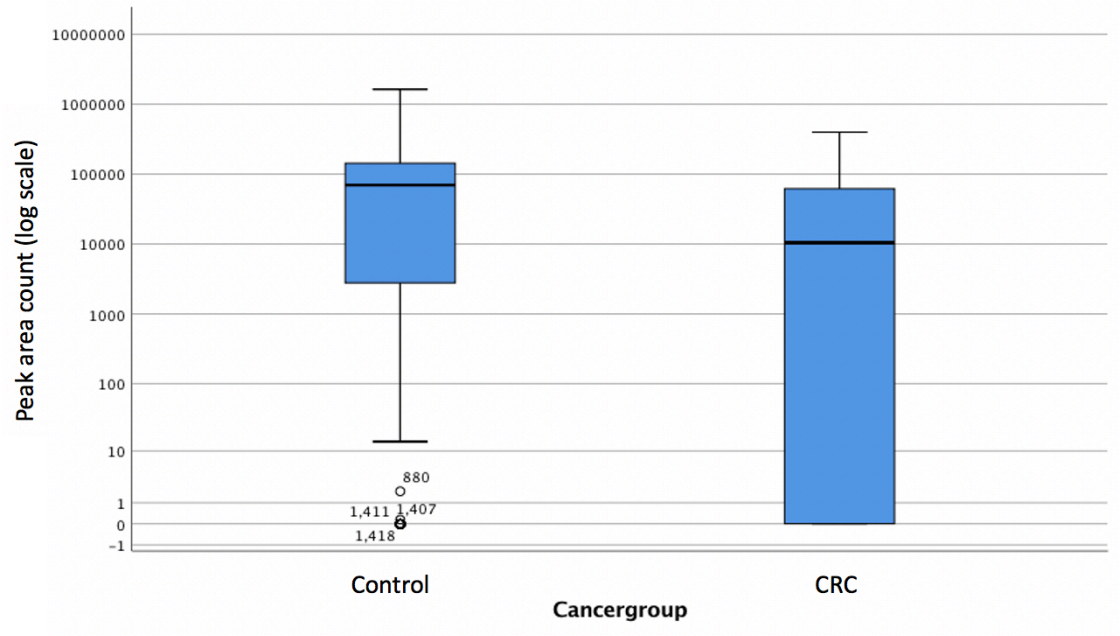


Figure 4-19 Boxplot of 3-ethyl-hexane abundance, VOC 15, in breath in non-CRC vs. CRC.

4.4.11 The alcohols

VOCs 4, 5, 10 and 13 were identified as 1,3-Dioxolane-2-methanol, 2-Phenoxy-ethanol, 2,2,4-Trimethyl-3-pentanol and 1-Undecanol respectively. These are all alcohols, and all had good matches with corresponding NIST library mass spectra, particularly in the case of VOC 4 and 5, making their tentative identities more confident. Appendix 13 shows the mass spectrum for each VOC and the comparative spectrum from the NIST database, to show how these compounds were identified.

VOCs 4 and 10 were found to be present in significantly higher abundance in the breath of patients with CRC (n=162) compared to those without CRC (n= 1270). VOCs 5 and 13 were found to be in lower abundance in CRC. The obtained peak area count for the two study groups are given in Table 4-11, and representative boxplots of the distributions in each group are demonstrated in Figure 4-20 to Figure 4-23. For each boxplot, the median is represented by the solid horizontal line, the whiskers represent the minimal and maximal value, and the box represents the interquartile range.

Table 4-11 Abundance, measured in peak area count, of four significant alcohols measured by GC-MS between patients with (n= 162) and without CRC (n=1270)

	VOC 4: 1,3-Dioxolane-2-methanol		VOC 5: 2-Phenoxy-ethanol		VOC 10: 2,2,4-Trimethyl-3-pentanol		VOC 13: 1-Undecanol	
	Non-CRC	**CRC**	**Non-CRC**	**CRC**	**Non-CRC**	**CRC**	**Non-CRC**	**CRC**
Median abundance	35.75	2665.74	1208066.0	670567.20	2593.60	5884.02	70160.40	26161.00
Minimum abundance	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Maximum abundance	70588.71	36149.09	37742018.00	36011833.00	759064.40	2013248.00	808136.20	367203.70
Standard deviation	6666.03	310168.72	2409058.00	3011138.00	61299.92	192091.70	92187.23	79499.80
Interquartile range	1450.47	8641.66	1444955.00	1002882.00	8450.10	19885.99	112391.70	92251.68
No. samples with zero peak area (percentage of total samples)	338 (23.6%)	18 (1.25%)	7 (0%)	1 (0%)	33 (2.3%)	3 (0.2%)	125 (8.7%)	47 (3.3%)

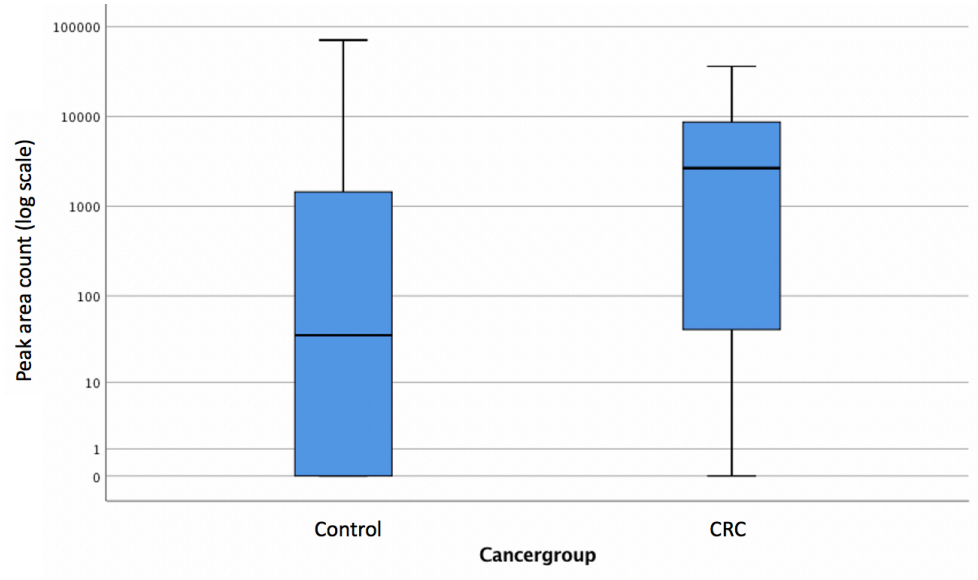


Figure 4-20 Boxplot of 1,3-Dioxolane-2-methanol abundance, VOC 4, in breath in non-CRC vs. CRC.

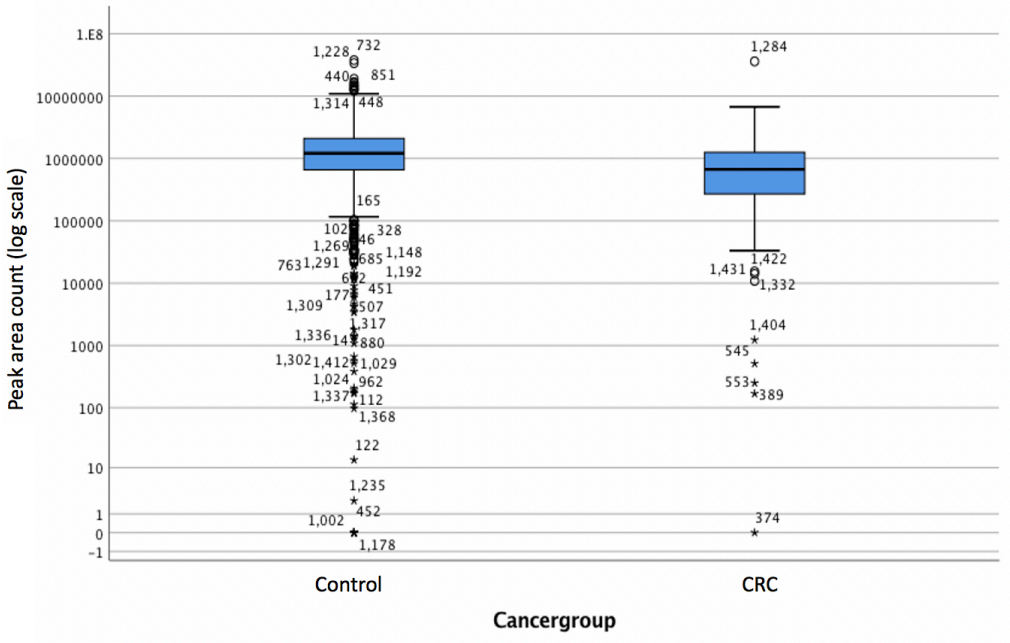


Figure 4-21 Boxplot of 2-phenoxy-ethanol abundance, VOC 5, in breath in non-CRC vs. CRC.

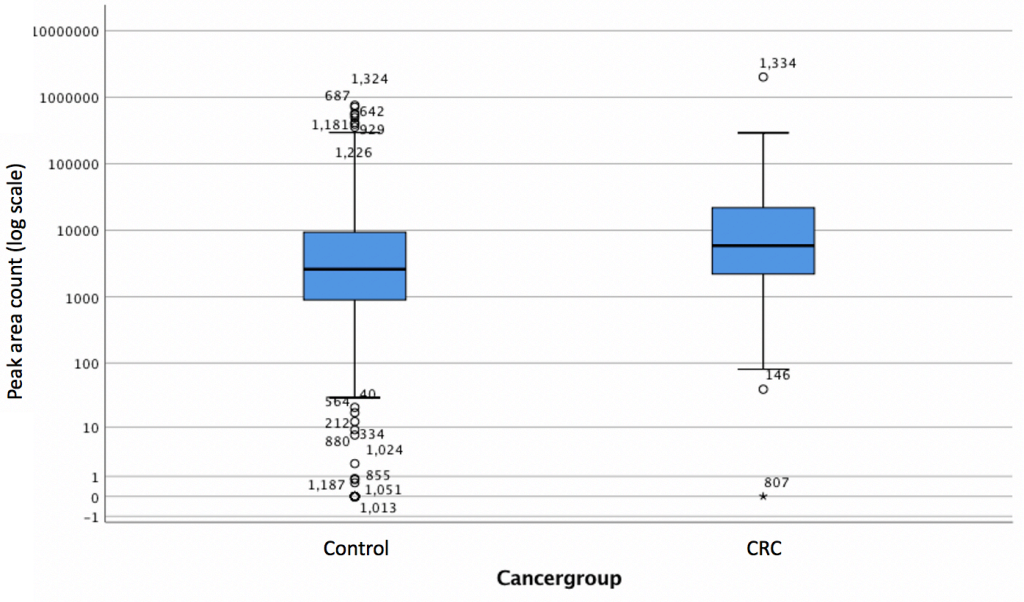


Figure 4-22 Boxplot of 2,2,4-trimethyl-3-pentanol abundance, VOC 10, in breath in non-CRC vs. CRC.

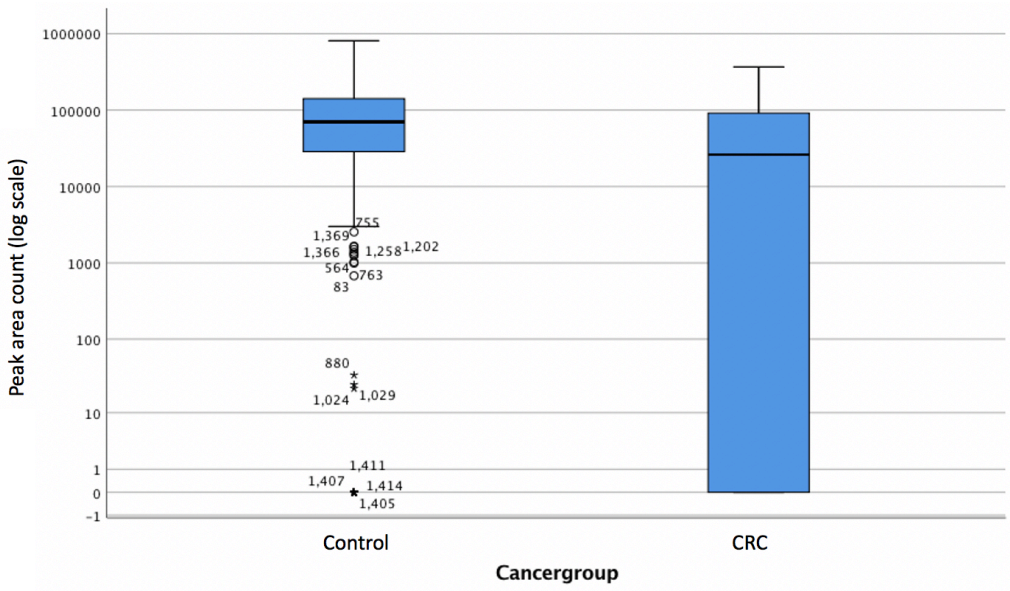


Figure 4-23 Boxplot of 1-undecanol abundance, VOC 13, in breath in non-CRC vs. CRC.

4.4.12 Phenol

VOC 14 was identified as phenol. Appendix 14 shows the obtained mass spectrum and the comparative spectrum for phenol from the NIST database, to show how it was identified. The spectral match was good, indicating confident identification. Phenol was found to be in lower abundance in CRC compared to controls. The obtained peak area count for the two study groups are given in Table 4-12 and the representative boxplot of the distributions in each group are demonstrated in Figure 4-24. The median is represented by the solid horizontal line, the whiskers represent the minimal and maximal value, and the box represents the interquartile range.

Table 4-12 Abundance, measured in peak area count, for phenol measured by GC-MS between patients with (n= 162) and without CRC (n=1270)

	VOC 14: Phenol	
	Non-CRC group	**CRC**
Median abundance	5611.56	916.61
Minimum abundance	0.00	0.00
Maximum abundance	340059.60	178120.90
Standard deviation	35943.55	19713.13
Interquartile range	14570.52	7057.58
Number of samples with zero peak area (percentage of total samples)	53 (3.7%)	15 (1.0%)

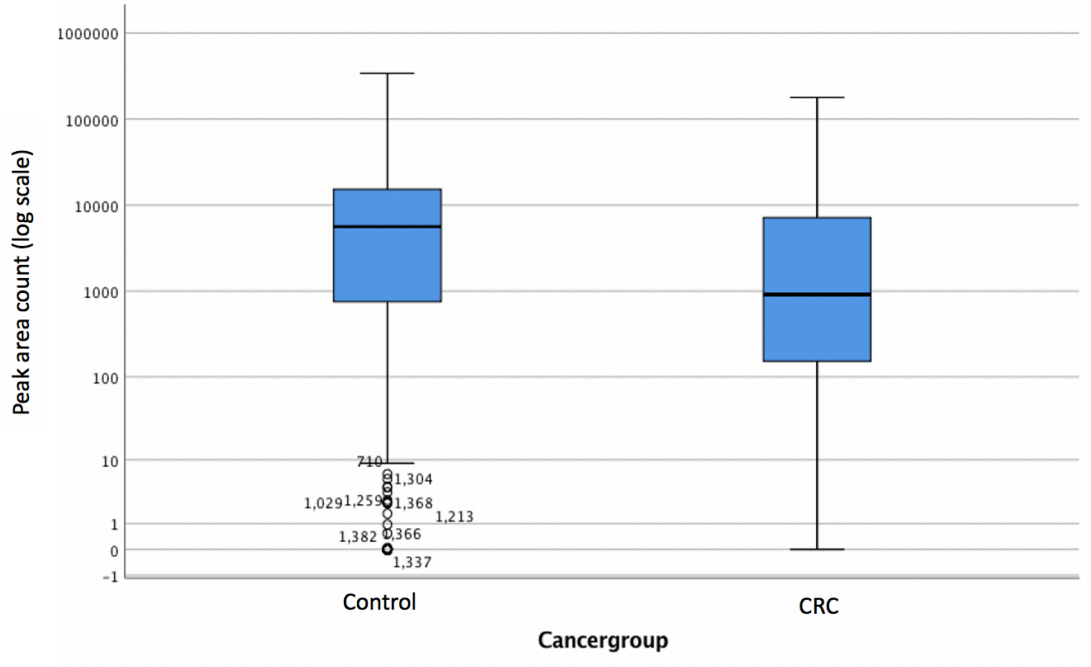


Figure 4-24 Boxplot of phenol abundance, VOC 14, in breath in non-CRC vs. CRC.

4.4.13 The non-aromatic cyclic hydrocarbons

VOCs 6 and 7 were identified as cyclopropane and 3,4-dimethyl-1,5-cyclooctadiene. Appendix 15 shows the obtained mass spectrum and the comparative spectrum for both from the NIST database, to show how they were identified. Both cyclopropane and 3,4-dimethyl-1,5-cyclooctadiene were found to be present in significantly higher abundance in the breath of patients with CRC (n=162) compared to those without CRC (n= 1270). The obtained peak area count for the two study groups are given in Table 4-13 and representative boxplots of the distributions in each group are demonstrated in Figure 4-25 and Figure 4-26. For each boxplot, the median is represented by the solid horizontal line, the whiskers represent the minimal and maximal value, and the box represents the interquartile range.

Table 4-13 Abundance, measured in peak area count, of two significant non-aromatic cyclic hydrocarbons measured by GC-MS between patients with (n= 162) and without CRC (n=1270)

	VOC 6:Cyclopropane		VOC 7: 3,4-dimethyl- 1,5-cyclooctadiene	
	Non-CRC	**CRC**	**Non-CRC**	**CRC**
Median abundance	11159.71	43148.59	32.59	469966.79
Minimum abundance	0.00	3.95	0.00	0.00
Maximum abundance	6903829.31	1025003.19	15932422.00	7568965.60
Standard deviation	524379.70	168532.40	719770.65	801986.31
Interquartile range	71442.29	147009.80	185631.97	988475.15
Number of samples with zero peak area (percentage of total samples)	22 (1.5%)	0 (0%)	631 (44.1%)	37 (2.6%)

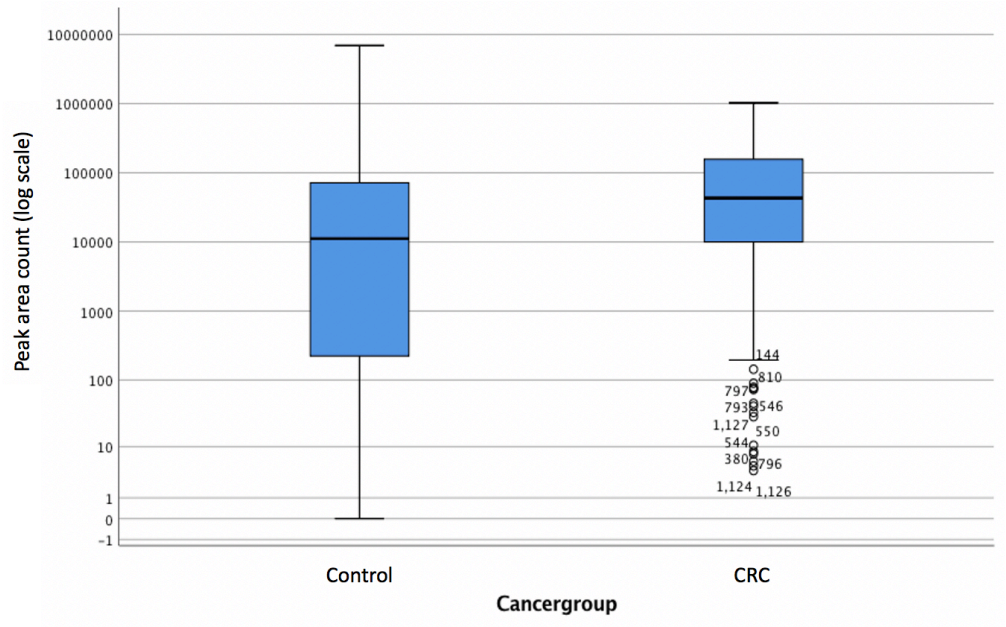


Figure 4-25 Boxplot of cyclopropane abundance, VOC 6, in breath in non-cancer vs. CRC.

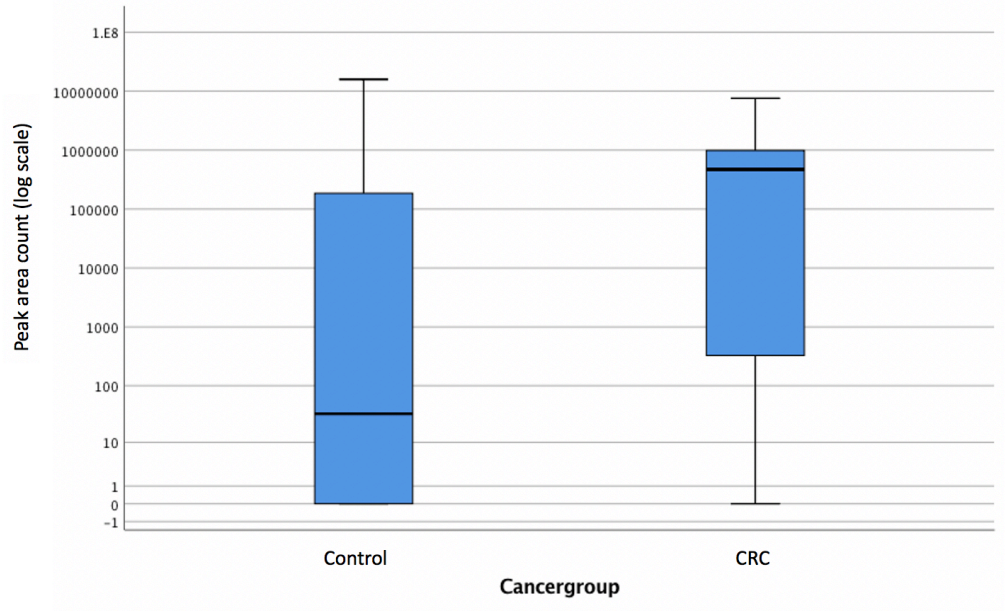


Figure 4-26 Boxplot of 3,4-dimethyl- 1,5-cyclooctadiene abundance, VOC 7, in breath in non-cancer vs. CRC.

4.4.14 Exploration of potential patient-related confounders

Age

Age was within the top 25 discriminating features between CRC and control patients in the machine learning analysis of the COBRA dataset. To investigate any effect of age on the predictive CRC VOCs found by this study, the machine learning prediction methodology was performed for different categories of age groups, in patients recruited between the start of the study and March 2019 (before the end of study recruitment). Analysis of age was done at this preliminary stage because it was important to define any effect of age on VOCs at an early point, in case the recruitment strategy needed to be altered to target/better represent a particular age group. Table 4-14 shows that the area under the ROC curve for prediction of younger cancer patients (40-59 years) was 0.89, compared to 0.78 for the oldest age group tested (70 to 89 years), however this was confounded by the lower number of CRC patients in the younger age group. When age groups were combined together for this preliminary dataset (n=1338), the sensitivity of the test was actually higher than for any specific age band, and the area under the ROC was 0.83 (Table 4-14). Comparison of the top 25 discriminating

VOCs for each age group band showed that the most discriminatory VOCs remained unchanged between age groups.

As a result of these findings, study recruitment was not altered; we continued recruiting all-comers and included patients of all age groups in the final analysis.

Table 4-14 Machine learning cancer prediction statistics for different age brackets, CRC vs. non-CRC patients

Age range (years)	Number of CRC patients	Number of non-CRC patients	Sensitivity	Specificity	Area under the ROC curve	ML prediction method
70-89	44	286	0.53	0.91	0.78	elastic
45-65	49	554	0.69	0.9	0.85	elastic
40-59	31	322	0.62	0.92	0.89	elastic
50-69	62	743	0.53	0.84	0.75	elastic
All ages	**120**	**1218**	**0.71**	**0.71**	**0.83**	**elastic**

Body mass index

BMI was found by the machine learning model to be linked to CRC diagnosis, see Table 4-6. This was found despite the demographic data of this study showing the same median BMI of 26 in both CRC and control groups, with no statistically significant difference on a Mann-Witney-U test, see Table 4-4. These results however only refer to the 51% of the 1432 included patients for which BMI was recorded at the recruitment stage. The result was also biased towards CRC patients (77.8% of CRC patients had a BMI recorded). It was therefore not possible to determine the effect of BMI on the CRC-specific VOCs found in this study because BMI was not adequately recorded in a large proportion of the patients sampled in endoscopy sites.

Fasting times

There was a statically significant difference in the overall fasting times between CRC and control patients, where CRC patients fasted for less time on average, see Table 4-4. Fasting time did not, however, feature in the top 25 discriminating features between CRC and control patients in the machine learning analysis of the COBRA dataset. Fasting time was therefore investigated for the specific 15 CRC VOCs identified by this study, using Pearson's correlation coefficient to compare fasting time and abundance of each VOC in question.

Table 4-15 below shows that for all of the 15 significant CRC VOCs found by this study, there was an extremely weak correlation between the VOC abundance and the number of hours fasted, near to zero in all cases except for potentially VOC 14 (phenol), where the Pearson's correlation coefficient was slightly higher at 0.176. These poor correlation findings indicate that, for the 15 significant CRC VOCs, fasting time was unlikely to be a confounder between the CRC and control groups, although it should be noted that 16% of the patients had missing data for fasting time.

Table 4-15 Correlation between significant VOC abundances (originally measured in peak area) with fasting time (hours), in 1197 patients with fasting data

Significant VOC number	Pearson's correlation coefficient between VOC abundance and fasting time in all patients
1	-0.098
2	-0.086
3	0.008
4	-0.093
5	0.048
6	0.097
7	-0.031
8	0.031
9	-0.025
10	0.010
11	0.022
12	0.027
13	0.044
14	0.176
15	-0.053

Effect of T stage on VOCs

The majority of the CRCs detected in this study (64%) were late stage cancers classified as T3 or T4 (see Figure 4-3 in section 4.4.3). To investigate the effect of T stage on VOC abundance, the top 15 discriminating VOCs were examined for all T stages by plotting median abundance on a line chart. Figure 4-27 shows that, without exception, all the VOCs that were found to be increased in CRC were found in higher abundance in T4 disease than in T1 disease. In the same vein, Figure 4-28 shows that all of the VOCs that were decreased in CRC were found in lower abundance in T4 disease compared to T1 disease, except for VOC 11, which was similar for all CRC T stages. As you can see from both figures, the magnitude of the change in VOC abundance varied for different VOCs, and the relationship to T stage was not linear for all VOCs, but this may be because the number of patients was not the same in each T stage group. The general trend however was maintained between T1 to T4 and the reverse, as already alluded to.

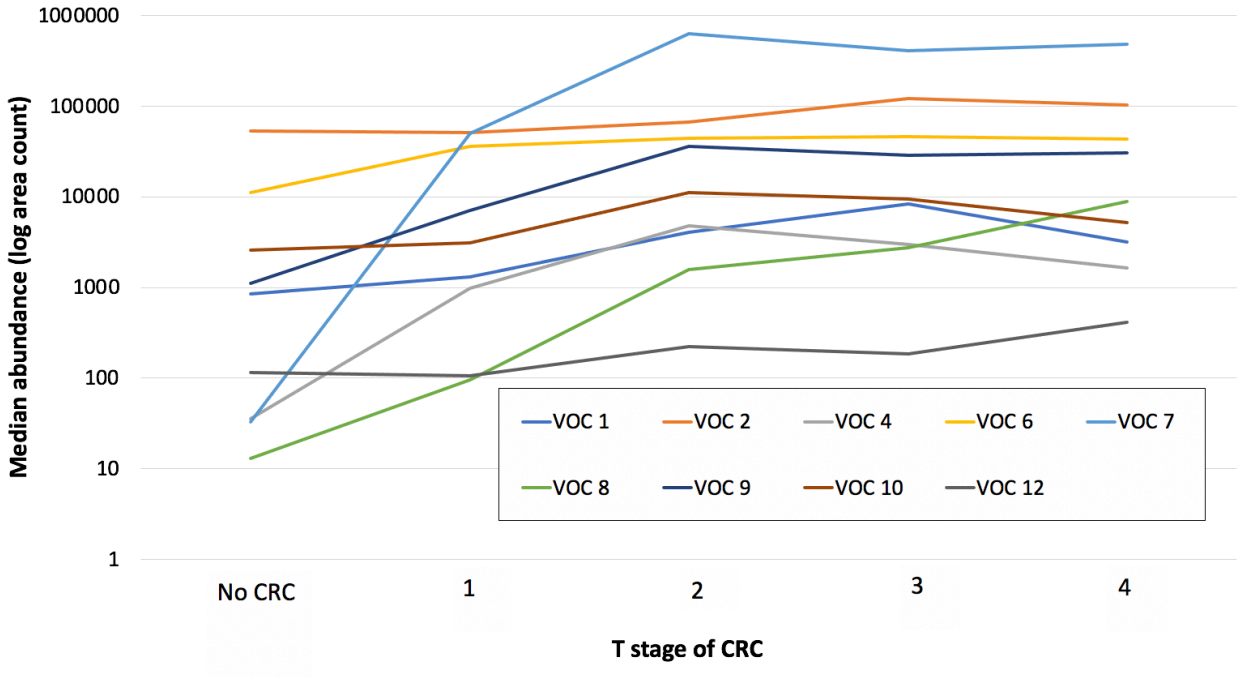


Figure 4-27 Median abundance of the 9 VOCs found to be significantly increased in CRC, by T stage at time of breath test, log scale.

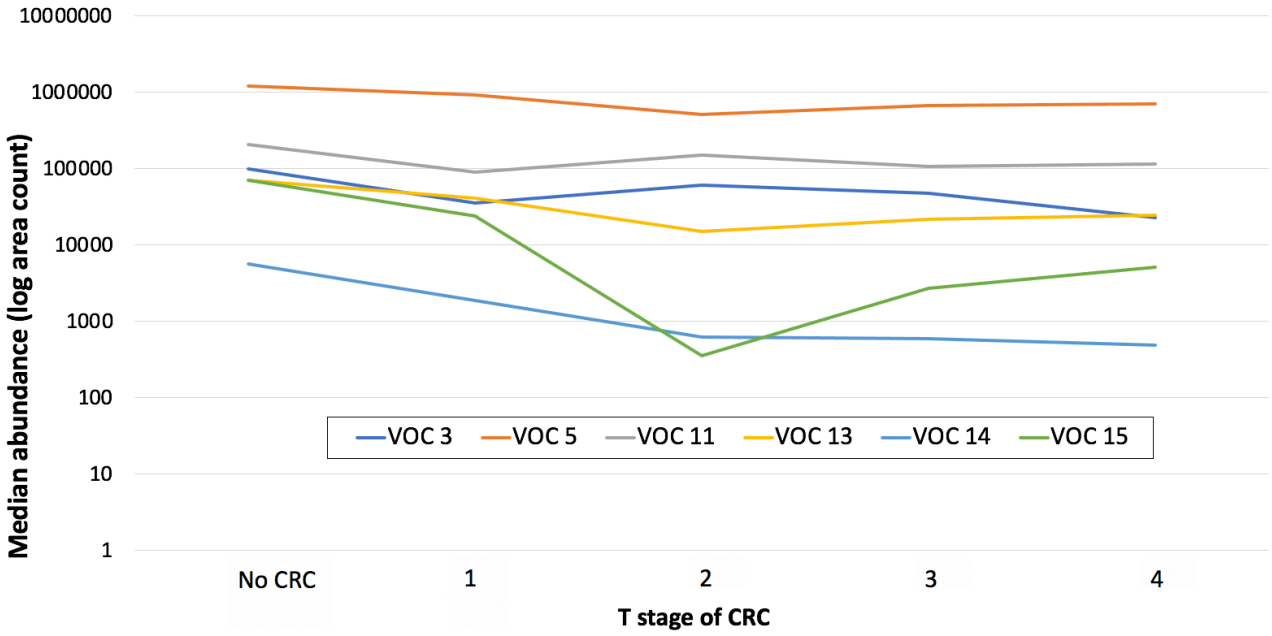


Figure 4-28 Median abundance of 6 VOCs found to be significantly decreased in CRC, by T stage of CRC at time of breath test, log scale.

Effect of site of tumour on VOCs

Most CRCs in the COBRA study were left sided (62%), see Figure 4-2, where rectum to splenic flexure tumours were classed as left sided for the purposes of this discussion. To investigate the effect of CRC site on VOC abundance, the abundance of the top 15 discriminating VOCs was compared for left and right sided tumours using a Mann Whitney U test. There were 100 left sided CRCs and 62 right sided CRCs. 13 out the 15 significant CRC VOCs showed no statistical difference in abundance between left and right sided tumours, where $p<0.05$. However, VOC 2 (dimethyl sulphide) and VOC 4 (1,3-Dioxolane-2-methanol) had p values of 0.002 and 0.048, respectively, when comparing tumour site (right vs. left sided), where the abundance of both was higher in the right-sided CRC group.

Effect of sampling environment on VOCs

As expected, the abundance of each of the top 15 CRC VOCs was significantly different when endoscopy and theatre sites were compared, where $p<0.01$, using a Mann Whitney U test. This was to be expected because the theatre patients were all CRC patients (with the exception of 2 patients), and the majority of endoscopy patients did not have CRC.

Endoscopy-diagnosed CRC (n=43) and theatre-diagnosed CRC (n=119) were therefore compared to create a pathology-matched comparison. Interestingly 9 of the 15 significant CRC VOCs were found to be significantly different between the theatre and endoscopy-diagnosed CRC groups where $p<0.01$; VOCs 1, 2, 4, 6, 7, 8, 10, 14 and 15. However, the CRC populations were markedly different from the two environments; theatre cancers were generally more advanced tumours than the cancers detected in the endoscopy unit, as the latter were largely asymptomatic patients who had been detected via the screening pathway (therefore with early cancers), see Figure 4-7. The higher number of theatre-diagnosed CRCs compared to endoscopy-diagnosed CRCs also made a comparison unequal. T stage of CRC was shown in the above section to have a marked, often linear, relationship to VOC abundance with regards to the 15 significant VOCs identified by this study, therefore effect of the sampling environment could not be measured from the COBRA data.

4.4.15 Results of machine learning prediction model, analysis 3 - colorectal polyps

The same machine learning analysis method was used to compare all detected VOCs from the high-risk, intermediate-risk, and low-risk polyp groups against those with no polyps, using GC-MS data. Different combinations and permutations of the 3 different polyp risk categories were compared with different combinations of patients with normal colonic findings, benign disease, IBD and CRC. As before, ROC curves were calculated based upon the results of a cross-validation method of 5 cycles of 5-fold stratified k-fold splits with reshuffling (explained in section 4.3.12). This means that, as before, the ROC curve was the average of ROC curves from individual runs, each with a slightly different feature selection and machine learning model. Up to 99 features were used to generate a given ROC curve, as determined by the machine learning algorithm, where “features” referred to individual ions but also individual clinical variables, i.e. any component that contributed to the separation of the groups. All comparisons and their results are detailed in Table 4-16. The ROC curves of the polyp comparisons with the best results, denoted by ** in Table 4-16, are shown in Figure 4-29 and Figure 4-30.

Table 4-16 Machine learning comparisons, done with different group combinations, to investigate the ability to predict colorectal polyps

ML analysis comparison groups	Explanation of comparator groups	Best Method (highest area under ROC curve)	Sensitivity	Specificity	Area under ROC curve
12_vs_456 (n=545 vs n=619)	Normal/benign vs any polyps	Elastic	0.62	0.57	0.63 +-0.01
123_vs_456 (n=619 vs n=592)*	Normal/benign/IBD vs any polyps	Elastic	0.66	0.54	0.65 +-0.01 **
12_vs_4 (n=520, n=337)*	Normal/benign vs LR polyps	Elastic	0.54	0.61	0.59 +-0.02
123_vs_4 (n=619, n=337)*	Normal/benign/IBD Vs LR polyps	Elastic	0.59	0.58	0.63 +-0.01
12_vs_56 (n=520, n=255)*	Normal/benign vs IR/HR polyps	Elastic	0.54	0.64	0.63 +-0.02
123_vs_56 (n=619 vs n=255)*	Normal/benign/IBD vs IR/HR polyps	Elastic	0.61	0.59	0.65 +- 0.01 **
12_vs_6 (n=545 vs n= 204|)	Normal/benign vs HR polyps	Elastic	0.49	0.68	0.62 +- 0.01
123_vs_6 (n=619 vs n=191)*	Normal/benign/IBD Vs HR polyps	Elastic	0.51	0.65	0.63 +- 0.02
1234_vs_56 (n=956 vs n=255)*	Normal/benign/IBD/LR polyps Vs IR/HR polyps	Elastic	0.45	0.68	0.59 +-0.01
1234_vs_6 (n= 956 vs n=191)*	Normal/benign/IBD/LR polyps Vs HR polyps	Elastic	0.43	0.68	0.58 +-0.03
12_vs_67 (n=520, n=311)*	Normal/benign Vs HR polyps/CRC	Elastic	0.61	0.68	0.70 +-0.02
123_vs_67 (n=619 vs n=311)*	Normal/benign/IBD Vs HR polyps/CRC	Elastic	0.63	0.68	0.72 +- 0.01
1234_vs_67 (n=956 vs n=311)*	Normal/benign/IBD/LR polyps Vs HR polyps/CRC	Elastic	0.54	0.69	0.66 +- 0.01
6_vs_7 (n=191, vs n=120)*	HR polyps Vs CRC	Random forest	0.68	0.91	0.85 +-0.01

ML denotes machine learning, HR denotes high-risk, IR denotes intermediate-risk, LR denotes low-risk.

*Some ML analyses were done before the study recruitment was fully complete, as a preliminary analysis for data exploration, presented here. This explains why numbers in the groups are slightly less than expected for some comparisons (depicted *). These analyses were not repeated using the full dataset because polyp group numbers were almost the same in both, and it would not have significantly altered the results.

** denotes the polyp comparisons that gave the highest area under the ROC curve result.

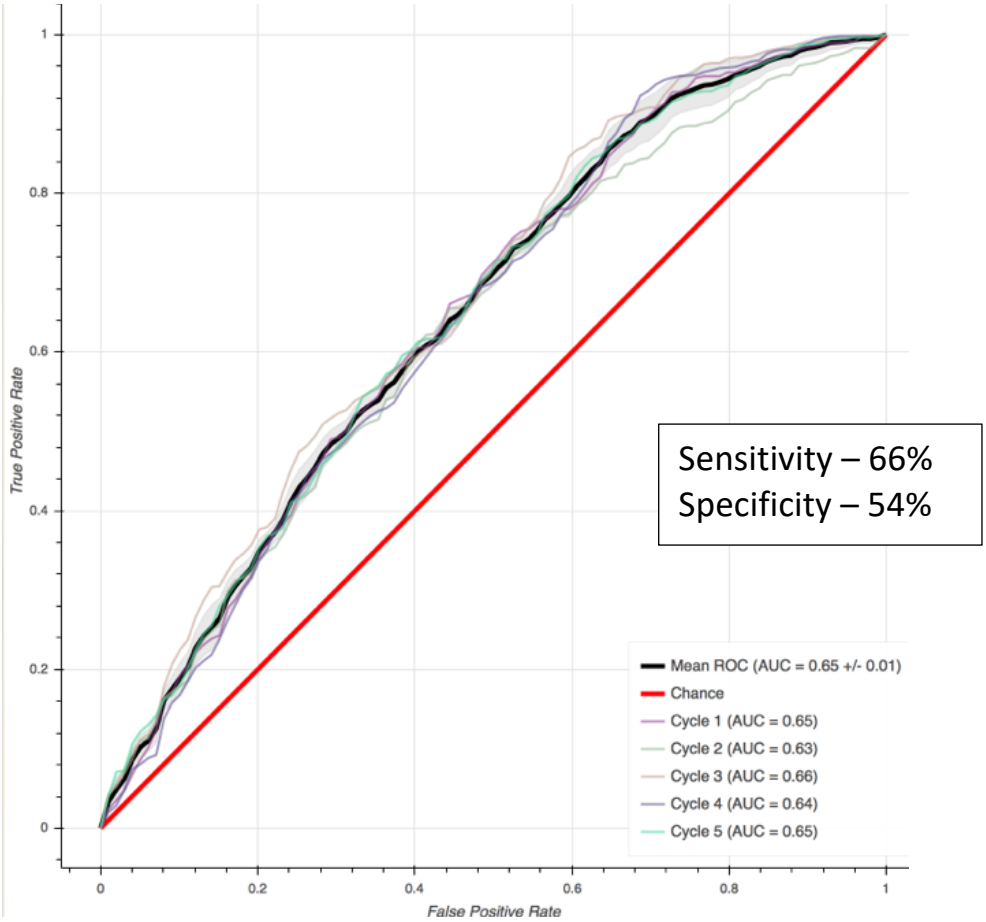


Figure 4-29 ROC curve of the prediction of normal/benign/IBD patients against all colorectal polyp patients (any polyp severity)

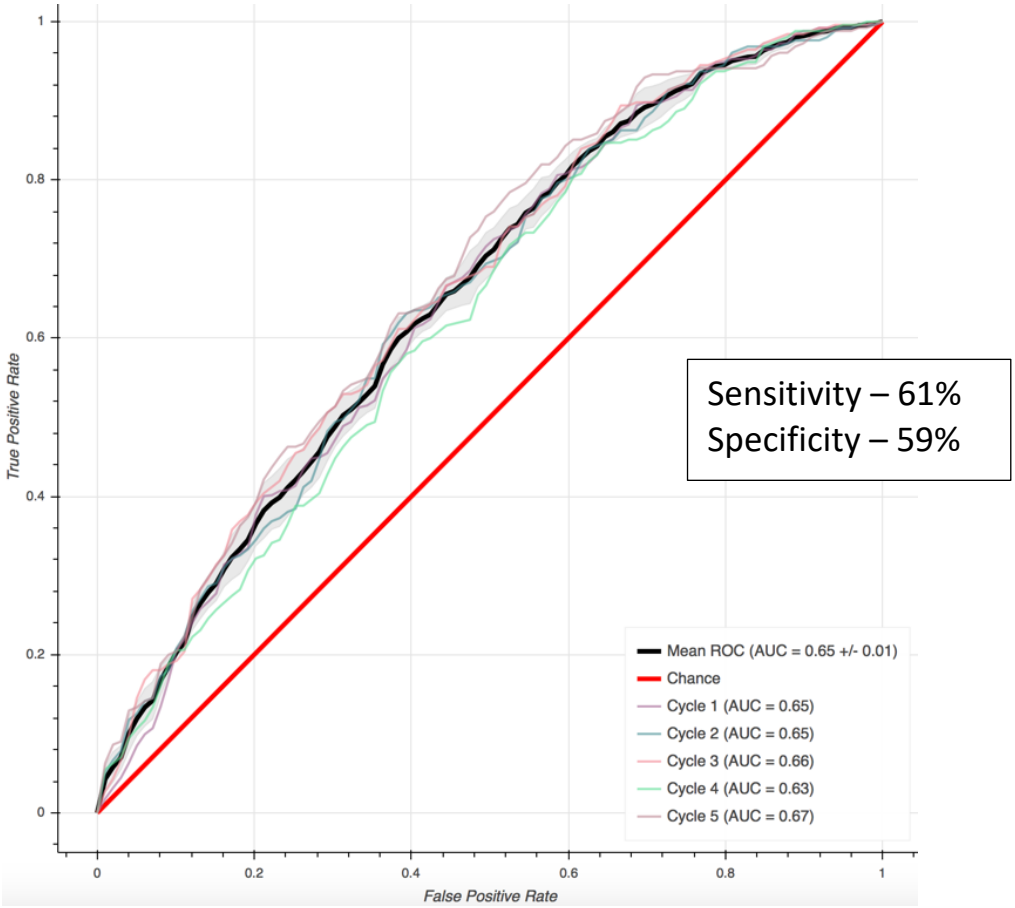


Figure 4-30 ROC curve of the prediction of normal/benign/IBD patients against intermediate-/high-risk polyp patients

The strongest model for the prediction of colorectal polyps of any risk category vs. no polyps tended to be the machine learning elastic method. The intermediate-risk polyp group was always combined with another group when used for analysis, as the patient number in the group was too small (n=67) to be compared in isolation. Table 4-16 shows the sensitivities and specificities when comparing different combinations of low-risk, intermediate-risk and high-risk polyp patients against normal and benign patients with or without IBD. The sensitivity of detection of any grade of polyps against positive and negative benign controls was 66% (+-0.05), specificity was 54% (+-0.06), the negative predictive value was 0.62 (+-0.02) and the accuracy was 0.60 (+-0.01). The area under the ROC curve was 0.65 (+-0.01), see Figure 4-29. When only intermediate-/high-risk polyps were compared, the sensitivity decreased to 61%, but specificity increased to 59%, still with an area under the ROC curve of 0.65 (+-0.01), see Figure 4-30. These findings show that the breath test was only partially able to detect the presence of polyps compared to patients with normal colonic findings, benign

disease of the colon or IBD, even where polyps were categorised as high-risk. When high-risk polyps were compared directly to CRC, the area under the ROC curve was 0.85, giving a result that was similar to the aforementioned CRC vs. non-CRC comparison reported in the previous section (which gave an area under the ROC curve of 0.87, see Figure 4-10). Where high-risk polyps were combined with CRC and compared to normal/benign patients, the results were inferior to that of the CRC group alone, meaning that the high-risk polyps diluted the ability to predict the CRC patients. The COBRA study demonstrated a moderate ability to detect colorectal polyps from breath using the ReCIVA breath testing tool, TD tubes and GC-MS breath analysis.

4.5 Discussion

The results of the COBRA study are very promising for detection of CRC from exhaled breath. CRC patients could be differentiated from positive and negative controls with a sensitivity of 77%, and a specificity of 87% and an area under the ROC curve of 0.87, see Figure 4-10. Polyps of any severity could be detected with a sensitivity of 66%, and a specificity of 54% an area under the ROC curve of 0.65, see Figure 4-29. Strengths of the study are the large number of patients, the multi-centre recruitment, and the blinded tube analysis. The data were analysed by an algorithmic learning platform for auto-deconvolution of GC-MS data to enable VOC molecular network analysis. (171, 172) There was also an integrated quality assurance system that employed SOPs for the entire breath analysis process to ensure the quality of breath sampling, transport and analysis in multi-centre clinical trials. The powerful machine learning prediction model used to analyse the breath data was capable of incorporating a wide range of clinical variables as well as over 1000 individual VOCs in the prediction model. For differentiating CRC from non-CRC patients, this meant that multiple possible clinical confounders were shown not to significantly influence cancer detection. A recent paper by our group provided a framework for conducting and reporting of studies that are investigating the role of VOCs for diagnosis of cancer. (102) This framework was used to guide the design of the COBRA study, aiming to standardise the processes for sampling and analysing VOCs in breath as much as possible, as well as the reporting of results.

The following discussion will first explore the individual CRC-discriminating VOCs that were found by the COBRA study, followed by the potential significance of these findings, followed by limitations of the study.

4.5.1 The esters

The propyl ester of propanoic acid (VOC 1), also known as propyl propionate, is an organic compound with the molecular formula $C_6H_{12}O_2$. It is a propanoate ester and results from the condensation of the hydroxy group of propanol with the carboxy group of propanoic acid. (177) It has a role as a human metabolite, a rat metabolite, and a fungal metabolite. (194) It occurs naturally in black elderberry, apple and cherries and is used a flavouring ingredient. (177) It is present in the breath of humans. (95) Propyl propionate was found by Mochalski *et al* to be one of the VOCs released by human hepatocellular carcinoma cells *in vitro*. (177) It was also found to be a specific discriminating marker in the headspace of stool of patients with active CD compared to healthy subjects. (124) Propyl propionate derives from propan-1-ol, which was found in the headspace of faeces collected from healthy volunteers and those with gastrointestinal diseases including IBD and gastrointestinal infection. (195)
VOCs 8 and 9 are allyl acetate and a similar overlapping ester possibly a branched version. The mass spectra are extremely similar, see appendix 10. VOC 12 is likely to be methyl 2-butynoate, although the comparative NIST mass spectrum does not show a perfect match. Ester concentrations in the headspace of faeces and urine have been linked to CRC (125, 132) as well as IBD (both CD and UC), non-alcoholic fatty liver disease (196) and coeliac disease. (197) (188) Esters in breath have also been found to be biomarkers of CRC. (117, 128) Whilst these were not the same esters that the COBRA study found, it is encouraging that similar compounds have previously been linked to cancers and colorectal disease, adding plausibility to our findings.

Esters in the body are produced by the condensation of the hydroxy group of alcohols with the carboxy group of a fatty acid. (177) Fatty acids are known to have a prominent presence in the gut lumen, due to fermentation of undigested carbohydrates by bacteria, particularly *Bacteroides.* (195) Gut bacteria have been shown to produce ethanoic, propionic, butanoic, pentanoic, and hexanoic acids. (198) Fatty acids are essential for maintaining health in the colonic mucosa, as they regulate epithelial homeostasis and have anti-inflammatory effects. (199) Several studies have shown that reduced levels of fatty acids in the gut are linked to gastrointestinal pathology including IBD. (200, 201) Branched chain fatty acids such as 2-methylbutanoic acid and methylpropionic acids have also be shown to be principally produced by the action of gut microbials on protein-rich foods (via the

respective branched amino acid). In particular, this occurs in bacteria such as *Clostridia spp*, and this has been shown *in vitro.* (202)

Esters are endogenously produced and known to be extensively metabolised in humans. For this reason, they could be potential candidate biomarkers. The human liver, small bowel and colon are known to contain an abundance of carboxylesterase enzymes. (203) These catalyse the cleavage of esters into the corresponding alcohol and carboxylic acid, where this could be part of endogenous ester metabolism. (204) It is already known that expression of the two most studied human carboxylesterase enzymes (CES1 and CES2) is markedly altered in cancerous tissue compared to healthy tissue. (205) It is possible that CRC cells disrupt carboxylesterase activity in some way, leading to higher levels of un-metabolised esters in the blood stream. The colon's venous drainage goes via the portal systemic circulation to the liver, for the purposes of metabolism of nutrients, it is therefore plausible that esters, alcohols and carboxylic acids could be detected in breath having been produced by a CRC cell then metabolised by the liver. It is also possible that alteration in the gut microbiome caused by CRC could alter levels of these related compounds (esters, alcohols and carboxylic acids) by alterations in production of fatty acids by bacteria in the gut lumen. For propyl propionate specifically, this hypothesis is supported by a study showing that propan-1-ol levels were found to be higher in tissue of human gastric cancer samples, as well as in the bacteria present in these patients, suggesting that the bacteria present in the stomachs of gastric cancer patients may cause the noted increase in propan-1-ol. (206) If this was the same in CRC, higher levels of propan-1-ol could lead to higher levels of propyl propionate (its direct metabolite). This gastric cancer study was limited however by small sample size (5 gastric cancer patients and 10 healthy controls). Further studies are required to test these mechanistic hypotheses.

4.5.2 Sulphur compounds

Dimethyl sulphide is a volatile sulphur-containing compound, predominantly produced by bacteria. (207) It has been found in multiple biological matrices, including blood, urine, tissue, stool and breath. (208) It is also present in plants and has a key role in the flavouring of beer and cooked vegetables. (209) Production in both humans and bacteria (anaerobic bacteria in the gut and mouth, as well as sulphate reducing bacteria in the gut) is ascribed to the metabolism of methionine and cysteine (sulphur-containing amino acids) as

part of the transamination pathway. (208) Dimethyl sulphide is a well-known breath VOC, found in the breath of healthy humans (95), and found to be dysregulated in patients with pathology states including liver disease (210) and hepatocellular carcinoma. (160) Importantly, it has also been associated with colorectal disease; Hicks *et al* found it to be one of the discriminatory compounds present in the breath of CD and UC patients, with an area under the ROC curve of 0.86 (when combined with a panel of 5 other VOCs). (114) It has also been found in faeces as a discriminator between high-risk polyp patients and controls in combination with a panel of 4 other VOCs. (112)

Related compounds dimethyl disulphide and dimethyl trisulphide have been previously reported in flatus (211), and in the headspace of stool from patients with a range of gastrointestinal diseases. (195) The related compounds hydrogen sulphide and methanethiol are also generated from sulphur-containing substances in the diet, where these compounds can be damaging to the large intestinal epithelium. (212) This is relevant to possible mechanisms of association with CRC, as one hypothesis could involve the causative link between the presence of volatile sulphur compounds in the gut lumen and CRC. Even if not causative, then this still is a plausible biomarker; high levels of dimethyl sulphide in the gut lumen could be caused by alterations in metabolism by CRC cells or by alteration of the microbiome as a result of CRC. Either of these mechanisms could explain high blood concentrations and therefore high breath concentrations of this compound. Further studies are required to test these mechanistic hypotheses.

4.5.3 The alkanes

Alkanes are readily found in breath and have been associated with both healthy and cancer patients in the literature, including 3-ethyl-hexane specifically. (95, 126) In a recent study of 40 lung cancer patients and 56 healthy controls using solid-phase microextraction (SPME)-aided GC-MS, differences in alkane concentrations were found to be the main discriminating feature between the cancer and healthy group, where 3 -ethyl hexane was one of the alkanes. (192) Alkanes have also been associated with gastrointestinal diseases. Dryahina *et al.* showed that pentane levels were significantly higher in IBD patients' breath compared to healthy controls, using SIFT-MS. This was postulated to be due to oxidative stress and lipid peroxidation that occurs in inflammation. (121) For CRC specifically, as highlighted by the literature review at the start of this thesis, many alkanes, methylated alkanes and

cycloalkanes were named as discriminatory markers for CRC. (84, 116, 117, 126-128). One study found 3-ethyl hexane to be associated with CRC using SPME-2D-GC-MS to analyse the plasma of 30 CRC patients and 30 controls. (126) 3-ethyl hexane concentrations in the plasma were significantly higher in the CRC group. (126)

Since they have been previously found to discriminate cancer patients, alkanes could be good candidate biomarkers for CRC. Possible mechanisms linking CRC and alkanes have already been postulated. One hypothesis is that aliphatic alkanes (and aldehydes) are present in cancer tissues as a result of oxidative stress leading to lipid peroxidation. Oxidative stress is well documented in the literature regarding its role in the development of cancerous mutations. (89) This is because reactive oxygen and nitrogen species cause DNA damage, leading to proliferation of mutations. (213) Specifically, when free radicals come into contact with the phospholipids that make up the cell membranes of human cells, oxidation occurs leading to formation of aldehydes, ketones, alkanes and carboxylic acids. (214) These products then react with other cellular components causing multiple effects such as altered membrane permeability, alteration in enzyme activity, altered ion transport and generation of metabolites, some of which are cytotoxic. (215) However, one flaw in this hypothesis is that branched alkanes may not be able to be formed by lipid peroxidation, because there are no branched polyunsaturated fats in the human body. (88) All of the alkanes discovered by the COBRA study were branched alkanes. Also, the abundance of the alkanes in question was lower in CRC patients than in control patients in the COBRA study, whereas the above mechanics suggest that they would be higher in CRC. However, this decreased alkane abundance phenomenon has in fact been found in a number of studies in the literature. (88, 145, 216) A study on lung cancer provided a mechanistic answer. Authors found that the alkane levels were lower in the breath of lung cancer patients (87), and postulated that this was because cytochrome p450 enzymes are induced during carcinogenesis, resulting in an increase in the rate of reduction of alkanes to alcohols. The alkanes that are not metabolised are 'excreted into the breath within minutes' because they have low solubility in blood. (88) (217) Abundance of alkanes in the blood would therefore be less than that of a person without the increased cytochrome p450 enzymatic activity. This could also explain the presence of the alcohols that were found in the breath of the COBRA CRC patients (see below section 4.5.4). There is also a link with microbial activity and branched alkanes. Bacteria are known to be able to metabolise branched chain alkanes, (218) where an altered gut microbiome could

potentially influence this metabolism cycle affecting the abundance of these particular alkanes.

4.5.4 The alcohols

1,3-Dioxolane-2-methanol (VOC 4), 2-Phenoxy-ethanol (VOC 5), 2,2,4-Trimethyl-3-pentanol (VOC 10) and 1-Undecanol (VOC 13) are all within the alcohol group. 1,3-dioxolane-2-methanol has previously been found in breath; and concentrations were significantly increased in the breath of 24 gastric cancer patients compared to 48 gastritis patients, measured with SPME GC-MS. (180) This was postulated by the authors to be due to the "Warburg" effect whereby glycolytic metabolism increases due to the increased proliferation characteristic of cancer cells. This leads to a shift towards anaerobic respiration of glucose, particularly in the stomach where even normal tissues are often under chronic anaerobic conditions. (180) 1,3-dioxolane-2-methanol could therefore be increased in gastric cancer as it is one of the products of anaerobic respiration. (180) This theory, if correct, could apply for CRC also. 2-phenoxy-ethanol has also been found in breath; it was present in 119 patients recruited for a chronic obstructive pulmonary disease study, measured using GC-MS. (182) It has not, however, been specifically associated with cancers in the literature. 1-Undecanol, a primary alcohol, is found naturally in many foods such as fruits, butter and eggs. It is used as a flavouring ingredient in food due to its floral citrus odour and fatty taste. (219) It has been found in urine and breath in humans. (188) A similar related compound, undecane, was one of the VOCs found by Altomare *et al* to be discriminatory for CRC in the breath of 48 CRC patients vs. 32 paired samples of the same patients post-curative CRC resection. (116) It was also found to be present in higher concentrations in the faeces of UC patients. (113)

It is plausible that the alcohols found by the COBRA study could be biomarkers for CRC because a wide variety of alcohols are endogenous and are found in the breath of humans. (95) The human liver and gastrointestinal tract contain alcohol dehydrogenase enzymes, (220) which are capable of reducing aldehydes to alcohols, (95) and also the converse reaction, oxidation of alcohols to aldehydes and ketones. (221) This is relevant because all VOCs produced from CRC pass into the blood stream via the gastrointestinal tract, and then circulate via the portal circulation to the liver. Disruption in different alcohol dehydrogenase enzymes could therefore account for both the increased and decreased levels of certain alcohols found by the COBRA study. Aldehydes and alkanes are thought to be present in

higher concentrations in cancer tissues as a result of oxidative stress and subsequent lipid peroxidation, as explained before in section 4.5.3. (89) Their presence could directly link to higher alcohol levels because, as explained before, cytochrome p450 enzymes are induced during carcinogenesis, resulting in an increase in the rate of reduction of alkanes to alcohols. (87) This could potentially lead to increased levels of alcohol in blood and therefore increased levels of alcohol in breath in CRC. However, decreased levels of alcohols could also be explained by increased alcohol dehydrogenase activity, controlling the interconversion between alcohols and aldehydes. However, one flaw in this hypothesis is that aldehydes were not found by the COBRA study to be significant compounds for CRC.

The levels of other alcohols in stool have also been shown to correlate with the presence of certain gut microorganisms, showing the relationship between alcohol levels and the microbiome. Ethanol concentrations in breath of patients with active CD were correlated with presence of *Bacteroides uniformis, Prevotella capri, Parabacteroides distasonis, Bifidobacterium ado-lescentis* and unidentified members of *Oscillospora, Coprococcus Blauta, Bilophilia, Ofdoribacter* and *Lactobacillus.* (222) The microbiome in CRC, and the resulting products of microbial metabolism, could significantly influence VOC levels in the blood and therefore breath.

4.5.5 Phenol

Phenol has been found in the breath, faeces, urine, skin and saliva of healthy humans. (95) In 2013, phenol was also found by our group to be one of 4 breath VOCs capable of predicting oesophagogastric cancer patients from positive controls, with an area under the ROC curve of 0.91, using SIFT-MS. (103) This was on the background of methyl-phenol having already been identified by our group as one of 7 VOCs that was significantly different between an oesophagogastric cancer and healthy control group, in gastric content headspace (also using SIFT-MS). (223) In a further study, looking at VOCs in urine, a lower level of phenol was found in the oesophagogastric cancer group compared to positive and negative control patients. (224) These previous findings of phenol in relation to cancer detection add weight to the findings of this study.

With regards to postulated mechanisms of the presence of phenol in the colorectal lumen, fermentation of tyrosine and tryptophan by gut bacteria in stool has been shown to produce the VOCs phenol and indole, respectively. It was previously postulated by our group

that increased phenol levels in the breath of oesophagogastric cancer patients could have been as a result of increased protein catabolism associated with cancer, changes in gut microbiota caused by gastrointestinal cancer (thereby altering the degree of tyrosine metabolism), alterations in the regulation of tyrosine metabolism caused by cancer cells, or a combination of all three of these. (104) These hypotheses could also hold true for CRC. With regards to the lower levels of phenol found in the CRC group in the COBRA study, this is the same finding that our group discovered when testing urine of oesophagogastric cancer patients against that of positive and negative control patients. (224) The group postulated at that time that lower concentrations of urinary phenol in the oesophagogastric cancer cohort may be due to dysfunctional phenol excretion secondary to alterations in tyrosine metabolism by gut bacteria or impairment in the removal mechanisms in blood. (224)

4.5.6 The non-aromatic cyclic hydrocarbons

Cyclopropane and 3,4-dimethyl-1,5-cyclooctadiene are both non-aromatic cyclic hydrocarbons. Cyclopropane has been documented in the breath and faeces of healthy humans. (95) Cyclopropane has also been found in higher concentrations in the breath of breast cancer patients compared to healthy controls. (183) Dienes were found in healthy humans by Costello *et al*, but it was felt that their origin may have arisen from smoking (particularly butadiene). (95) However, Filipiak *et al* performed a study looking at VOCs in the breath of smokers vs. non-smokers, and 3,4-dimethyl- 1,5-cyclooctadiene did not appear in the list. (225) In the COBRA data, the levels of 3,4-dimethyl- 1,5-cyclooctadiene were not significantly different between current smokers (n=185) and non-smokers (n=1223), p=0.087 on Mann-Whitney-U test. Cyclopropane is potentially a plausible biomarker for CRC for the same reasons as documented above for alkanes (cyclopropane is a cycloalkane). 1,5-cyclooctadiene has been documented as a naturally occurring compound; it has been found in the plant *A. thaliana* using high-resolution mass spectrometry and carbon-13 isotope labelling. (184) It is therefore potentially plausible as a biomarker for CRC.

4.5.7 Potential significance of the COBRA findings

The COBRA study has yielded promising results regarding the ability to detect CRC and colorectal polyps using breath. In order to contextualise these results, it is useful to discuss them in the context of currently used CRC and adenoma screening and detection tools,

namely the FIT test. For screening of asymptomatic patients, at a threshold for positivity of >50 μg/g (the English threshold for positivity for screening is currently 120μg/g faeces), the sensitivity of the FIT has been reported as 67% and specificity as 96% for detection of CRC. (226) More recently, at a threshold for positivity of >20 μg/g, the sensitivity and specificity of the FIT has been reported as 71% and 95% respectively for detection of CRC in screening of asymptomatic patients at average risk of CRC. (42) The breath test results have demonstrated superior sensitivities than FIT when used at these thresholds. However, the cohort for this current COBRA study only has 46% of patients recruited via the BCSP, so it is not an equal comparison. Also this study did not test the breath of the BCSP patients without a positive gFOBT/FIT, who were therefore not invited for colonoscopy. This study therefore keeps open the possibility that a breath test could be used for screening for CRC, but is inconclusive, as this study did not assess a representative screening cohort. With regards to symptomatic patients, the sensitivity of FIT for detecting CRC in patients with lower risk symptoms (as per DG30 criteria NICE guidelines) (22) has been reported as very high at 92.1-100%, but with a lower specificity than the breath test of 76.6-85.8%, using the OC-sensor or HM-JACKarc FIT. (25, 227-231) However, again the COBRA cohort was not the same patient group, as only 59% of the COBRA patients were in the symptomatic group at the time of their colonoscopy, and not necessarily with low-risk symptoms. A recent study of FIT in a mixed cohort of symptomatic patients, fulfilling either the NICE NG12 (20) high-risk for CRC or the NICE DG30 (22) lower-risk for CRC criteria, was done in 619 patients at a single centre. (23) This showed sensitivities for FIT of 84.8 to 100%, with specificities of 81.3 to 91.8%, using the low threshold for positivity of 10 μg/g. In comparison to these findings, the breath test performed well, with comparable specificity albeit lower sensitivity. However, if the ROC curve for the COBRA breath test is used to demonstrate what the corresponding sensitivity of the test would be if the specificity was 81% (the same specificity reported for the FIT), then the corresponding sensitivity of the breath test would be 81%. This figure is more comparable to the FIT result for a high and lower CRC risk mixed cohort, giving promise to the hope of breath testing being a useful predictive tool for CRC.

For polyp detection, there is currently little data on the sensitivity and specificity of the FIT. A study of patients with abdominal symptoms reported on the accuracy of the FIT for the detection of high-risk adenomas and CRC combined, using a low threshold for pick-up of 10 μg/g. The sensitivity was 59 - 63% and specificity was 81 to 98%. (25) In patients having

surveillance after a previous polypectomy (originally a screening cohort), even at the low threshold of 10 μg/g, the FIT was only able to detect advanced adenomas with a sensitivity of 56.6% and specificity of 73.7%. (232) The breath test, with its sensitivity to detect any grade of polyp of 66%, exceeds that reported in either of the above two studies (albeit in a mixed asymptomatic and symptomatic cohort), although specificity of the breath test was lower at 54%. However, in order to compare the FIT against breath testing, a dedicated comparative study would need to be performed. This could be planned in future work.

As an additional point, regardless of which test is superior for CRC detection, the breath test may be more acceptable to patients and therefore have improved uptake compared to the FIT. In general practice, stool-based tests are known to represent a compliance challenge. (233) It is also known that in the screening population, almost half of the eligible population do not return a stool sample. (234) The MAGIC study, presented in chapter 5 of this thesis, shows the universally high acceptability for breath testing in primary care. If high acceptability were to indeed lead to high uptake of the breath test, then this may give it an important place in the arsenal of tools for CRC detection. The breath test could also potentially be used in combination with the FIT, if this was shown to give superior predictive results. This is a plausible potential option as even with both tests combined this would likely be cheaper per patient than doing a colonoscopy, whilst still avoiding an invasive procedure in many. In summary, the breath test has shown promising results, making it worthy of further investigation and direct comparison to the FIT.

4.5.8 Discussion of potential patient related confounders

For the CRC vs. non-CRC breath comparison using the machine learning prediction model, whilst many clinical confounders were not shown to be predictive, age and BMI were in the top 25 features differentiating the groups, see Table 4-6.

Age

Co-linearity between cancer state and age is well documented, where increasing age is a risk factor for CRC. (235) The CRC group was significantly older than the control group, as shown in Table 4-4 (66.5 years vs. 63 years, respectively, $p<0.001$). It is therefore perhaps unsurprising that age was found by the machine learning model to be a predictive factor for CRC, even though separate machine learning analyses also showed that the best predictive

results actually came from the all-age cohort rather than any specific age band, see section 4.4.14. Age is therefore unlikely to be a major confounder in this data. Future studies should ensure that a wide variety of age bands are recruited, and make efforts to match the control group for age.

BMI

Correlation between cancer stage and BMI is also well documented, where increasing BMI is a risk factor for development of CRC. (235) However, late stage cancers of all types can also lead to low BMI due to the increased catabolism associated with severe illness. With regards to breath specifically, BMI is known to affect breath acetone levels (92, 236), where levels are strictly linked to food intake and energy balance. Whilst neither acetone nor any other ketone was in the top 15 cancer differentiating compounds found by the COBRA study, BMI still needs to be considered as a potential confounder. A future study should include the specific measurement of BMI as part of the breath testing protocol rather than as a box to fill in on the patient data sheet. This may encourage more complete recording of BMI data than was done in this study.

Fasting times

Fasting time is potentially a confounder in the data because fasting state is known to affect VOCs, namely acetone (92, 236). The effect of fasting on the abundance of cancer-specific VOCs found by this study is not documented in the literature, but fasting time was found to have a poor correlation with VOC abundance when the COBRA data was investigated, see Table 4-15 in section 4.4.14. The poor correlation reported in the results section indicate that, for the 15 significant CRC VOCs, fasting time was unlikely to be a confounder between the CRC and control groups, although it should be noted that 16% of the patients had missing data for fasting time. In a future study, greater efforts should be made to ensure that this data is always recorded on the study information sheet, as this information is recorded without exception for patients attending hospitals for any procedure.

Bowel preparation

Bowel preparation could be a possible confounder with regards to VOC detection, as it by definition changes the contents of the colon, influencing the contribution of bowel contents and likely the microbiome to the VOC profile. Differentiation of patients before and after bowel preparation was shown in a study using e-nose sensors, showing an effect of bowel preparation on breath, but without VOC identification or knowledge of causative factors. (96) Another small study used GC-MS to analyses the breath of 61 gastric cancer patients before and after bowel preparation prior to a colonoscopy. (237) They found that only acetone was significantly elevated after bowel preparation, but that this could have been a consequence of fasting (acetone is well known to increase in fasting). (238) Bowel preparation did not appear in the list of top discriminating variables for differentiating CRC and non-CRC patients in the COBRA machine learning analysis. This was interesting given that a variety of bowel preparations were used in the study, and some of the theatre-recruited patients received no bowel preparation at all, resulting in a statistically significant difference in distribution of bowel preparation used between CRC and control groups, $p<0.001$, see Table 4-4. Future studies should look specifically at the effect of bowel preparation on breath VOCs, and ideally give the same bowel preparation to all patients.

Other comorbidities

It is likely that the presence of other pathological conditions, including inflammatory or infective states, could influence VOC production. (239, 240) Effect of medications could also influence VOCs. (241) Diagnoses of previous CRC or heart disease were significantly higher in the CRC population than in the control patients, $p<0.001$ and $p=0.009$, respectively, see Table 4-4. The significantly higher use of warfarin in the CRC group is likely linked to the increase in heart disease, and the use of laxatives and antibiotics is likely to be due to bowel symptoms associated with their CRC diagnosis. Prevalence of comorbidities such as high blood pressure, diabetes and other cancers were equal between groups. It is difficult to tease apart the effect of the CRC from the effect of heart disease on the top 15 VOCs in breath, also limited by the fact that 'heart disease' was a non-specific term that included ischaemic heart disease but also cardiac arrhythmias. Future studies should have more detailed and specific past medical history recorded, particularly regarding cardiac pathology, given the results of this study.

Effect of T stage on VOCs

This study showed that abundance of all 15 of the CRC-specific VOCs was related to the T stage of the tumour, often with a linear relationship. These findings support the specificity of the breath VOCs identified by this study, as very likely a more invasive (larger) tumour could alter metabolism to a greater degree, and therefore have a greater influence on breath content. This is a very exciting finding that adds weight and plausibility to the findings of this study and the conclusion that breath can be used to detect CRC. The correlation with T stage and VOC abundance may also explain why polyps were not detected very well by the breath test, as the order of magnitude of the alteration of breath VOCs may have been significantly smaller than in T1 tumours, if the same trend followed.

Effect of site of tumour on VOCs

The results of this study showed that the abundance of dimethyl sulphide and 1,3-dioxolane-2-methanol was significantly higher in the right-sided CRC group than in those who had left-sided tumours. It is not clear why the abundance of these compounds would vary depending on site of CRC. If this finding were shown to be consistent in future studies, it could potentially be extremely clinically useful; higher magnitudes of change in dimethyl sulphide and 1,3-dioxolane-2-methanol could signify need for a colonoscopy, rather than a flexible sigmoidoscopy.

Effect of sampling environment on VOCs

As documented by Schubert *et al*, high concentrations of an inhaled trace gas in the ambient environment may disproportionately affect the alveolar concentration gradient of the compound in question, and therefore affect elimination in the exhaled breath. (242) There is no consensus as to how to correct for potential effect of inspired trace VOCs. Previous work done by the Imperial VOC group has shown that some trace gas concentrations (hydrogen cyanide, isoprene, acetic acid and ammonia) were equivalent in room air samples from different environments, including a hospital theatre. However, some trace gases (propanol and ethanol) varied in concentration depending on the environment. (243) Another more recent study by the group measured room air in theatres, endoscopy and clinic environments over 5 days, at different time points in the day, and found that the majority of trace VOCs were present in equivalent concentrations in each of these areas with the

exception of butanal, acrolein, butanol, pentanol, butyric acid, putrescine, methanol and isoprene. VOCs measured included a wide range of fatty acids and aldehydes as well as phenol, methyl phenol and ethyl phenol. (163) A further study done by the Imperial VOC group (for oesophagogastric cancer diagnosis) analysed room air and compared it to breath for potential cancer-differentiating compounds, including phenol. The atmospheric phenol measured from the surgical sciences laboratory was present at a negligible concentration compared to the concentration from patients' breath samples, where the highest single room air reading was still lower than the lowest readings from any of the human samples (healthy, positive controls and oesophagogastric cancer patients). (104) Room air was therefore thought to have negligible impact on the cancer differentiating VOCs for this study. In this COBRA study, efforts were made to eliminate the effect of trace VOCs in ambient air by the use of the 'Clean Air Supply Pump for ReCIVA '(CASPER) filtration system, as explained in chapter 3 of this thesis. This is a filtration system whereby air is entrained through a charcoal column to rid it of volatiles before inhalation via the ReCIVA face mask. This is a well-recognised method of mitigating for the potential problem of room air contamination. (243) However, there is contention about the period of time required of breathing cleaned air to allow the successful washout of previously inhaled contaminant VOCs; some studies suggest 4 minutes, whereas other suggest longer periods. (217, 244) The ReCIVA breath sampler has a brief 'learning' phase at the start of a sample where the patient breathes cleaned air without breath collection onto TD tubes. This is however significantly less than the 4 minutes of 'washout' suggested in the literature. In addition to the use of the CASPER filtration system, environmental variability was minimised by replicating the setting as much as possible over the sampling sites. All patients were sampled seated in a room away from the procedure room, prior to entry, before any exposure to sedation, throat spray or anaesthesia. The sampling process and sample handling procedures were the same across sites. Also, four different endoscopy units were used, where the lowest percentage contribution that any endoscopy site had to recruitment was 13%. There was therefore a significant number (>200) of patients from each endoscopy site, hence any environmental factors unique to one site should not have been able to overwhelm the data. Three different theatre sites were also used with a recruitment contribution of 2-5% of the data each. Despite these attempts to standardise sampling, future studies should eliminate any question of environmental influence on breath VOCs by sampling room air at the time of every patient sample.

4.5.9 Sampling considerations and limitations

The combination of the ReCIVA® breath sampler with TD tubes provided a sampling solution that allowed collection of breath at 7 different sites over London, with the ability to safely transport and store VOCs until mass spectrometry analysis. This was the only commercially available sampling option for such a large study. However, there remained some limitations:

TD tubes

Storage times and stability of VOCs on the carbograph/tenax sorbent within the TD tube has already been touched upon in chapter 3 of this thesis. TD tubes have been shown to be able to safely store VOCs without significant degradation or contamination for one month at room temperature, and 1.5 months at -80 degrees Celsius. (138, 139) In this study the maximum storage time of TD tubes at room temperature from sampling to GC-MS analysis was 8 days (in unfrozen tubes), well within the recommended storage time. Tubes that were frozen were stored for longer (median 17 days, with a standard deviation of 40 days) and were analysed on the same day as they were thawed. Our laboratory QC processes minimised to the best of our ability any variability between TD tubes, but it is true that as the sorbent ages it may have become contaminated with heavy/sticky VOCs resistant to cleaning. It is also true that not every tube will necessarily behave identically when desorbed. Future studies could examine specific target VOCs and ensure that the chosen sorbent is optimal for the specific chemical properties of the VOC in question.

ReCIVA® breath sampler

Previously ascertained optimised sampling parameters for ReCIVA® breath collection were used for this study (volume, flow rate and breath fraction used). (143) These parameters however do not change the fact that the ReCIVA® breath sampler relied on two pumps to draw the exhaled breath from the face mask through the TD tubes, at a rate of 200mls/min, until 500mls breath had been sampled. It is possible that these pumps could have varied in flow rate throughout the breath sample. This very problem was reported for the ReCIVA® sampler in a recent paper, where this was suggested as a possible cause for variability in the repeatability and accuracy of VOC measurement. (144) Regular manual calibration was

advised. Flow rates were checked manually every 4 weeks during COBRA, and often small variations were found between pumps and between repetitions. As a pragmatic practical solution in a busy clinical study, an average flow rate was therefore used to judge if the unit had passed the QC or not. Small variations could however influence VOC concentrations, and prevent the 4 TD tubes from being considered as replicates. The ReCIVA® also has the ability to collect mouth or nose air, or both, dependent on how the patient chooses to breathe into the mask. All patients were asked to breath normally whilst wearing the mask during COBRA, where the fraction of breath collected was 'whole breath'. Mouth breath has been shown to contain higher levels of VOCs that originate from the oral cavity, namely ammonia, ethanol and hydrogen cyanide. (245) These VOCs were not among the discriminatory CRC VOCs discovered during this study, but contamination from the oral cavity and careful consideration of breath fraction measured is none the less a consideration for future work.

4.5.10 Gas chromatography-mass spectrometry considerations and limitations

Despite being the gold standard for VOC analysis, capable of detecting a wide range of VOCs at a low limit of detection, the GC-MS has some limitations as an analytical platform. Firstly, it requires ions to be thermally stable and volatile in order for them to be separated and then detected. In the analysis of VOCs, volatility is not a concern, but some thermally unstable compounds may go undetected. Secondly, it is not inherently quantitative, unless a calibration curve has been developed using a standard for the compound in question. This is a particular issue in untargeted biomarker studies such as COBRA, because the cancer-differentiating compounds were unknown until the end of the study. Obtaining a standard and then creating a calibration curve 2 years later, even if the same GC-MS column was used, may give inaccurate quantification because the column may drift over time. Baseline drift can occur over time due to gradual damage to the column due to heating and cooling, as well as exposure to oxygen or chemical damage from inorganic acids or bases, on exposure to non-volatile materials. (246) Retention time shift can also occur due to increases/decreases in column temperature or gas flow rate, any leaks in the injector or contamination of the column. (246) For these reasons a calibration curve must be developed over the same time period as the comparative analysis. This lack of quantification meant that the 15 significant VOCs could not have a threshold for positivity calculated. A third limitation of the GC-MS is time for analysis, which was 50 minutes per sample in our lab. This affected the study

because, ideally, two or more breath sample replicates should have been analysed per patient, particularly as the ReCIVA® tool is capable of sampling onto 4 TD tubes simultaneously. This was not done due to throughput limitations, as there was not enough available instrument time alongside all the other studies being performed by our lab. A fourth limitation was the column used. This is a very important discussion point for the results of the COBRA study.

4.5.11 Gas chromatography mass spectrometry column choice

GC-MS columns have different affinities for polar and non-polar compounds, affecting the ability to separate and detect these. The column used for the COBRA study was the Zebron ZB-624 capillary column (60m x 0.25mm ID x 1.40 µm df; Phenomenex Inc, Torrance, USA). This column is designed for analysis of VOCs, suggested for environmental VOCs by the manufacturer, as it is a mid-polar column, good for universal detection. It can therefore detect alkanes (non-polar), aldehydes and ketones (mid-polar) very effectively. However, highly polar compounds such as short chain fatty acids (SCFAs) are unlikely to be detected by the ZB-524 column. This is an important limitation because fatty acids have been found in the literature to be associated with cancer, including from the Imperial VOC group, where acetic, butyric, pentanoic, and hexanoic acids were used to predict oesophageal cancer. (86, 163) Studies finding SCFAs as cancer markers have used analytical platforms such as the SIFT-MS, where SIFT-MS does not rely upon a column as it has no chromatography phase. SIFT-MS can pick up compounds that are ionised by reagent ions, including SCFAs. The SIFT-MS however would not be able to detect alkanes as these are inert and don't react with the reagent ions. There are therefore pitfalls with each of the analytical platform choices, and where the GC-MS is chosen, the column choice limits coverage of detected volatiles. This explains some of the differences in findings between studies, including those previously done by our group.

4.5.12 Study design considerations and limitations

COBRA was a multi-centre cohort study, the largest in the literature to date, assessing VOCs for the detection of CRC and polyps. Patients entered the study having been referred for endoscopy or theatre via different diagnostic pathways. The majority of endoscopy patients were referred as asymptomatic patients as part of the BCSP. Whilst this was a very effective way to recruit patient with polyps (63% of BCSP patients in this study had polyps

found), this proved to be an inadequate route to reach our expected prevalence of 8% CRC diagnoses within our target recruitment number, as only 4.5% of BCSP patients were found to have CRC in our London cohort. Sampling in theatres was therefore initiated to provide an enriched CRC cohort. This introduced confounding factors with regards to the CRC T-stage and the environment of breath testing, because theatre patients tended to be symptomatic and have a later-stage cancer diagnoses than the largely asymptomatic patients in the endoscopy CRC group, as discussed earlier. Interestingly, of the 2WW referrals that were sampled, the yield of cancer was 5.2% and polyp yield was 29%; however, only 96 patients were sampled from the 2WW pathway. In a future study, the cohort sampled would need to be considered carefully. A likely better way of recruiting CRC patients would be to sample only 2WW referrals, as cancer yield would likely be higher than in the BCSP, and in a large enough study this may avoid the need to supplement CRC numbers by sampling in theatre (which introduced a breath-testing environment confounder between sites). This 2WW patient population may also more accurately represent the future cohort to which a breath test may be offered, particularly as the COBRA study found that polyps could not be reliably detected by breath, limiting its potential to be a cancer screening and prevention tool.

A further design limitation was the unavoidable incompleteness of the blinding process. For the patients sampled in endoscopy, the study was blinded with no knowledge of patient diagnosis at the time of the breath test, by either patient or operator or indeed endoscopist. Patients sampled in theatre with known CRC could not have their diagnosis blinded, representing a limitation in the study. However, researchers analysing the breath samples in the laboratory were blinded to the diagnosis of the patients, as clinical details were only paired with VOC analysis results at the very end of the study.

Another possible design limitation that could have contributed to the relative inability of the COBRA study to reliably detect patients with polyps could be the way that the polyp patients were grouped. The high-risk polyp categorisation for this study was based upon British Society of Gastroenterology (BSG) guidelines from 2002 for risk of CRC development, where a high number of polyps and/or a polyp greater or equal to 1cm attracts more frequent colonoscopic surveillance due to increased chance of CRC development. (8, 174) Adenomas with villous or high grade dysplastic features were also put in the same category of ‘high-risk’, on advice from experts in the field, in advance of an update to the 2002 BSG guideline which was to include high-grade dysplasia (published in 2020), and international definitions that

include villous histology as high-risk. (7, 247, 248) This customised and adapted risk profile was used to categorise polyp patients for the COBRA study. With regards to alteration in breath VOCs however, it seems plausible that even a very high number of very small adenomas may disrupt the mucosa or alter the microbiome less than a large or villous or dysplastic adenoma. There are no studies looking at this. If this was the case, and if breath could be used to detect large/villous/dysplastic adenomas (theory untested by COBRA data), this test would anyway have limited use for detection of early cancer, as it would miss the 'high-risk' of CRC patients who had multiple but small polyps.

A further limitation of the study design was the lack of repeatability of VOC measurements by the same analytical platform. The ReCIVA® breath sampler measured four TD tubes simultaneously for each breath sample in this study, but only one of the TD tubes was analysed by GC-MS, mainly due to analytical instrument time availability. The other three TD tubes were analysed using three different ionisations using PTR-ToF-MS, a faster analysis platform. This was used to demonstrate capacity for high-throughput. However, whilst the obtained PTR-ToF-MS data in future could be used to corroborate and correlate to the GC-MS findings, the analytical software and techniques for establishing the identity of VOCs using PTR-ToF-MS were still in development and were not easily translatable to the GC-MS findings at the time of analysing these results. This could be part of the future work, to provide more robust methods of aligning and correlating PTR-ToF-MS and GC-MS datasets. In a future study, as recommended by a recent Imperial publication on challenges in breath research, "advanced analytical techniques may be required for better sensitivity using TOF-MS; further separation of target VOCs with GC-GC and accurate chemical identification using MS-MS". (102) If this was used, this would negate the need to analyse with different analytical platforms, thereby avoiding the introduction of complicated data comparison difficulties later. In a similar vein, a further limitation of the study design was lack of demonstration of reproducibility. One Agilent GC-MS was used to analyse all samples, at the central laboratory at St Mary's Hospital London. Ideally, in future multi-centre studies, the same analytical platform should be used at each of the laboratories, to demonstrate that the findings are the same regardless of analysis site.

4.5.13 Limitations caused by data loss and exclusion

Data loss limited the numbers of patients that could be put through to the statistical data analysis phase, and would have significantly weakened the power of the study if this problem had not been almost entirely compensated for by intentional over-recruitment. The recruitment target of 1463 was exceeded by 26.8% (1855 recruited), yet still the actual number of patients successfully analysed by the GC-MS, with good quality data, used for statistical analysis, was 1432. The study was therefore underpowered by 31 patients. This was not thought to be significantly detrimental to the analysis however, because the target recruitment of 117 CRC patients was exceeded (162 CRC patients were included in the final analysis). However, the overall data loss was higher than expected. As presented in the recruitment section of the results, data were lost due to patient exclusions, human error, laboratory instrumental breakdown, but the most significant cause for data loss was failure of samples to pass the required QC standards. This accounted for 55% of data loss, where approximately a third of these patients were excluded due to ReCIVA®-related software or quality problems, and two thirds were excluded because of inadequate acetone levels on the TD tube, indicating inadequate/unreliable breath collection. In future studies, efforts should be made to improve the reliability and repeatability of sampling methods so that less data is excluded and valuable time and resources are not wasted. This could include using different sampling tools or techniques.

4.5.14 Human factors

Human factors can significantly affect the collection and transfer of breath samples to the laboratory, as well as influence the meticulousness of adherence to QC protocols. (102) Chapter 5 of this thesis, the Breath MAGIC study, investigates in detail the feasibility and acceptability of breath sampling in a primary care setting. Primary care is a realistic setting for a future breath test. Future studies should investigate the human factor element of breath testing in secondary care, as COBRA did not investigate this specifically.

4.5.15 Establishing breath analysis as a diagnostic or monitoring tool

Significant challenges towards establishing breath analysis as a diagnostic or monitoring tool have been summarised succinctly by Filipiak *et al* previously. (82) The first challenge is the need for standardised processes for sampling and analysing VOCs in breath.

The importance of this in terms of QC processes has been addressed in chapter 3 of this thesis, and the wider issue was also explored in depth by a recent paper by the Imperial group (102), which was used as the framework for design and reporting of this study. Secondly, technical challenges must be considered. These include the fact that ions are trace compounds in breath, therefore a significant change in a disease state may in reality be represented by a concentration difference of only a few ppb. This is in the context of including abundant ions in the same analysis, hence effects of less abundant ions may be dwarfed. This is also complicated by the fact that a single ion is not necessarily associated with a disease state, but more likely a panel of ions. For the initial untargeted analysis that is required for biomarker discovery, powerful statistical modelling is required to process many thousands of permutations and combinations of potentially predictive ions. The model must also be able to incorporate the numerous clinical variables, any of which could potentially confound the data, and be able to account for these. Filipiak *et al*, go on to highlight the need for unambiguous identification of biomarkers, which must have a plausible biochemical explanation for their generation, as well as an explanation for why they are altered in a disease state. This is in the context of variability between and within patients, caused by genetic and environmental factors, as well as timing of sampling and presence of co-morbidities. All of these considerations are relevant to the COBRA study, and may limit the reliability of the findings.

A further challenge to implementation is logistics and cost. The cost of the current experimental equipment would be prohibitively high for use in a wide-scale NHS setting, particularly in the context of a comparable triage tool such as the FIT, which costs approximately £4 - £9 per test. (24) However as discussed before in this thesis, the ReCIVA® and TD tubes both have limitations, which for this study resulted in a significant amount of breath data exclusion. It is therefore unlikely that a future study or triage tool for clinical practice would use the same equipment and set-up. Improved sampling options and procedures are already in development both locally, and likely globally. Future cost-saving options could also include the development of sensor technology for detection of targeted CRC compounds, which would need to have been quantified (for determination of diagnostic thresholds) and validated in a further study of breath testing for CRC (see future work section).

In order for a breath test to be implemented as a diagnostic tool, the patient population it would serve needs to be determined, as well as the benefit it would bring to current practice. This would help ascertain its place in a clinical pathway. Both screening and urgent referral pathways have been in place for over a decade for CRC; UK guidelines for urgent referral perform well for the minority of CRC patients with high-risk symptoms but perform less well for the majority of cases with low-risk symptoms. This has led to the introduction of the FIT for low-risk but symptomatic patients as part of the NICE DG30 guideline. (22, 249, 250) Improving the diagnostic pathway for low-risk symptomatic patients is important, as 24% of CRCs are diagnosed through GP referrals via routine pathways and experience substantial delays as a result of the concentration of resources on urgent referrals. (47, 251, 252) This could be where the breath test would be most useful. As discussed earlier, the breath test had comparable results to that of the FIT, with higher sensitivities for a screening population than the FIT, and comparable specificities in the symptomatic population. Further work needs to be done to look at the breath test's performance in specific symptom-groups of patients (as the COBRA cohort was mixed symptomatic and asymptomatic patients), ideally with direct comparison of the FIT and the breath test. This could also include investigation into the use of both FIT and breath testing combined, as the sensitivity and specificity data could be perhaps improved if these two platforms were used together. Even with two tools used per patient this would likely still represent a cost saving in comparison to a colonoscopy being performed.

4.5.16 Future work

The immediate next step is the COBRA 2 study. This is planned to be a 5700-patient blinded multi-centre clinical study of all patients attending for an urgent 2WW pathway colonoscopy. More sophisticated sample analysis will be performed using 2-dimensional chromatography (GC-GC) for compound separation and time of flight mass spectrometry (TOF-MS) for detection and quantification, in 2 separate GC-TOF-MS replica instruments. This would allow analysis of the breath samples from each patient using both a polar and non-polar column, optimising capture of VOCs from different chemical classes. Quantification of discriminatory VOCs would enable establishment of a test threshold for separating patients with cancer from controls (both positive and negative controls).

A more thorough QC process will also be adhered to, having learnt lessons from the current COBRA study, including daily sampling of room air at all sites. This would also incorporate more rigorous attention to TD tube cleaning and storage protocols to ensure minimal contamination of TD sorbents was possible. This would include running the TD tubes on the PTR-TOF-MS after conditioning, both to desorb them once more but also to act as a check for residual VOCs. After conditioning, TD tubes would be stored in an oven with nitrogen flowing through them until they were to be used for sampling. All breath tested patients will also perform a FIT test, allowing direct comparison of the two tests, as well as allowing a predictive model to be built based on results of both tests combined.

We anticipate that the results of COBRA 2 will be stronger than those from the current COBRA study due to the superior analytical technique proposed, use of polar and non-polar columns, routine use of calibration curves for the significant CRC VOCs to more accurately quantify them, and more robust QC procedures including regular measurement of room air at all sites.

4.6 **Conclusion**

The COBRA multi-centre study showed that exhaled breath could be used to detect CRC of any stage from positive and negative controls with a sensitivity of 77%, specificity of 87% and negative predictive value of 97%, in 1432 patients attending hospital for a colonoscopy or for CRC resection in theatre. The area under the ROC curve was 0.87. Polyps of all severities could be detected with a sensitivity of 66% and specificity of 54%, with an area under the ROC curve of 0.65. The use of the ReCIVA breath sampler and TD tubes allowed sampling to occur concurrently over 7 London hospitals, with feasible storage and transportation of breath samples, maintaining QC standards for the breath sample in the majority of cases. The GC-MS analysis technique was able to process high numbers of samples, without the need for breath pre-concentration, with a low detection limit able to measure trace VOCs at sub-ppb levels. Confident identification of significant VOCs was performed using comparison of mass spectra with the NIST library. The study was supported by an algorithmic machine learning platform for auto-deconvolution of GC-MS data to enable VOC molecular network analysis. (171, 172) 15 VOCs were extracted from the list of possible CRC biomarkers, included dimethyl sulphide, phenol, and compounds from the ester, alcohol, alkane and non-aromatic cyclic hydrocarbon chemical classes. All had highly plausible mechanistic hypotheses that could explain their role

in CRC. These 15 VOCs together were able to predict the presence of CRC from positive and negative controls using breath with an area under the ROC curve of 0.83.

These breath test results provide the basis for a larger multi-centre trial, moving a step closer to the implementation of this innovative and highly acceptable tool for reliable and non-invasive CRC and polyp detection into clinical practice.

4.7 **Funding**

COBRA was awarded a special purpose fund from Rosetree's Trust and Stoneygate Trust via Imperial College Health Charity (grant number 171812a), which paid for consumables and courier costs. It was also supported by infrastructure funding from the NIHR London In Vitro Diagnostic Co-operative programme (grant number P67522). Work conducted by NIHR affiliated research nurses was funded as part of their usual salary by the NIHR by virtue of the fact that COBRA was adopted into the NIHR research portfolio.

5 CHAPTER 5: THE METHODOLOGICAL APPROACHES TOWARDS A GASTROINTESTINAL CANCER (MAGIC) BREATH TESTING STUDY, BASED IN PRIMARY CARE

5.1 **Background**

So far this thesis has highlighted the burden of disease of CRC and the need for new biomarkers to aid early diagnosis and streamline existing referral pathways (chapter 1). It has also reviewed the literature on the use of VOCs for detection of different colorectal diseases (chapter 2), and provided encouraging results showing that a breath test may be able to detect CRC and polyps with a high sensitivity and specificity (chapter 4 - COBRA study). This chapter introduces the idea that a breath test could be used for multiple different gastrointestinal cancers, with their own specific VOC biomarkers. This makes sense as different gastrointestinal cancers may share some common symptoms. This chapter also bases the breath test in primary care, as this is likely the most useful place for a diagnostic tool, before invasive/expensive investigations are reached.

Gastrointestinal cancers (including CRC) are a major cause of morbidity and mortality. Oesophagogastric cancer is rising in incidence, now representing 7% of all gastrointestinal malignancies internationally. (253-255) It has a very poor prognosis with only 10-30% survival at 5 years. (256, 257) If detected early this can translate into an improvement in survival of up to 80% at 5 years. (258) Symptoms are non-specific, such as dyspepsia and acid reflux, therefore diagnosis may be difficult, especially as upper gastrointestinal symptoms make up 3% of all consultations in primary care. However, investigating all patients with upper gastrointestinal symptoms using endoscopy (the gold standard test) would be time and resource intensive, and therefore not feasible in the UK NHS. (259) Pancreatic cancer has a very poor survival rate, 3.3% at 5 years, and often has silent or non-specific symptoms. (260) Only 10% of pancreatic cancers are treated with curative intent. (261) There is no diagnostic screening test for pancreatic cancer, and patients are referred straight for CT scan, involving considerable doses of radiation. In order to improve the diagnosis of gastrointestinal cancers, efforts should be made to determine how the risk of cancer can be assessed using sensitive primary care tests that would need to be valid, easy to perform, and well tolerated by patients.

Encouraging work has been done on the diagnostic accuracy of VOCs for the detection of gastrointestinal cancers by groups including our own at Imperial College London. A recent multi-centre validation study of 335 patients showed a sensitivity of 80% and specificity of 81% for the diagnosis of oesophagogastric cancer using a panel of 5 VOCs in breath. (163) A recent prospective pilot study followed by a validation study (already alluded to in chapters

2 and 4) showed that propanal in breath had a sensitivity of 83.3% and specificity of 84.7% for detection of CRC. (109) Both studies used SIFT-MS to analyse VOCs in breath. This is in addition to the more recent COBRA study presented in chapter 4 of this thesis.

Despite these promising findings from the literature, there is currently no consensus on how a breath test could best fit into current cancer diagnostic pathways, or whether it would be feasible and acceptable to perform it in a primary care setting. It has been recognised in the literature that there is a great need for this. (102) This study therefore aims to contribute to currently running parallel diagnostic studies by investigating some of these questions.

5.2 Aims of study

Primary aim: To inform the design of future large-scale breath testing studies based in primary care by investigating the feasibility of different recruitment and sampling strategies.

Secondary aim: To examine the acceptability of the breath test from the perspective of patients and GPs.

5.3 Ethical considerations

The ethical approval for the study protocol and supporting documents was obtained from the London - Camden & Kings Cross Research Ethics Committee (14/LO/1136) prior to the study starting in December 2016. Sponsor and clinical commissioning group (CCG) approval was obtained and the study schedule of events and statement of activities was approved by the Health Research Authority. All clinical study officers (CSOs)/nurses/research staff had up to date Good Clinical Practice certification and appropriate training in research methodology, as well study-specific training. The study was conducted in accordance with the recommendations for physicians involved in research on human subjects adopted by the 18th World Medical Assembly, Helsinki 1964 and later revisions.

5.4 Methodology

The MAGIC (Methodological Approaches Towards a Gastrointestinal Cancer) breath testing study aimed to test the breath of 1000 patients attending to see their GP, in one of 26 participating GP practices in London over a 12 month active recruitment period, spanning 18 months (details of GP practices are in appendix 16). The recruitment period was split up into

Phase-1 (29/11/16 to 26/5/17- 6 months) and Phase-2 (7/11/17 to 14/6/18- 6 months) in order to test different sampling and recruitment models, explained later.

5.4.1 Eligibility

Study eligibility criteria were patients aged 18 to 90 years old who were suffering from upper or lower gastrointestinal symptoms. Gastrointestinal symptoms included all 2WW and urgent referral symptoms within NICE guidelines. (20, 262) GPs and trial staff were provided with a list of all eligible gastrointestinal symptoms (appendix 17-19). Patients with persistent symptoms (lasting >2 months) were included only if they had ongoing requirement for pharmacological control. Patient eligibility was assessed by GPs at the time of a routine face-to-face appointment or from review of electronic medical records.

5.4.2 Breath Sampling

Participants were consented to perform a breath test either by NIHR nurses or CSOs (part of the central study team) or trained GP practice nurses local to the particular GP practice. NIHR nurses/CSOs were allocated to the study via the support of the northwest London clinical research network. CSOs/nurses explained the breath test procedure to participants, provided them with a participant information leaflet (see appendix 20), and once they had read it asked them to sign a consent form (appendix 21). They then collected the breath sample.

The breath test was performed using a CE-marked handheld breath sampling (ReCIVA®) device (Owlstone, Medical Ltd, Cambridge, UK), used as per the manufacturer's recommended sampling procedure, at standard optimised settings previously published (143), and explained in this thesis, see chapter 3. This same breath sampling methodology was also used for the COBRA study, chapter 4 of this thesis. Essentially, a sterile rubber facemask (single use) was fitted onto the ReCIVA® handheld device and the participant was asked to perform normal tidal breathing whilst wearing the mask until a defined volume (500mls) of breath had been sampled. The device had a cable attachment to a small laptop, whereby the computer software controlled the settings, recorded quality parameters, and indicated when the full volume of breath had been collected. During sampling, exhaled breath was entrained using inbuilt pumps through a TD tube (Markes International, Llantrisant, UK) packed with Carbograph/Tenax sorbent, designed to absorb and retain the

VOCs present in breath. Sampling took around 5 minutes on average, as breath was drawn by internal pumps present in the ReCIVA® handset at a flow rate of 200mls/minute, where pumps only turned on during exhalation as triggered by rising carbon dioxide levels in the mask. In order to standardise the inhaled air that the patient breathed prior to each exhalation, the facemask had an inlet for inflow of room air that had been passed through an activated charcoal filtration column. This limited any effect of possible room-air contamination.

For each participant the CSO/nurse filled out a clinical details form, detailing past medical history, medications and key information such as smoking status and timing of last oral intake. Clinical details forms and TD tubes used for that individual's breath sample were linked by labelling with the same unique study ID. This was anonymous and unidentifiable, with no ability to retrospectively trace the true patient identity. In the latter part of the study, during Phase-2 of recruitment, CSOs/nurses also asked patients to fill out an anonymous patient acceptability questionnaire (appendix 22). At the end of a sampling period at any given GP practice (mostly two-week blocks), CSOs/nurses asked GPs from the participating practice to fill out a questionnaire about their perspective of the breath test (appendix 23).

5.4.3 Sample processing and breath analysis

After sampling, the TD tube was removed from the ReCIVA® breath testing device, capped and transported back to the laboratory at St Mary's Hospital London for analysis. Samples were transported by courier or by a member of the research team at the end of each sampling week. Samples underwent analysis via mass spectrometry at St Mary's Hospital London. Two types of mass spectrometry were used; GC-MS and PTR-ToF-MS, introduced in chapter 3 of this thesis. The PTR-ToF-MS (model 1000, Ionicon Analytik GmbH, Innsbruck, Austria) was run using one of three different reagent ions H_3O^+, NO^+, O_2^+, coupled with a TD customised autosampler unit (TD100-xr, Markes Ltd, Llantrisant UK). The PTR-ToF-MS method for VOC analysis is extensively described elsewhere. (142) The GC–MS (Agilent 7890B GC with 5977A MSD (Agilent Technologies, Cheshire, UK)) was also coupled with a Markes TD-100 (Markes Ltd, Llantrisant UK) TD unit, with a previously developed analytical method. (169)

Breath VOCs sampled on TD tubes were analysed as soon as possible from when they arrived in the lab. PTR-ToF-MS data regarding breath chemical content were extracted using

PTR-MS Viewer version 3.2.8.0 (Ionicon Analytik) and analysed using in-house generated scripts written with R programming language. (166) GC-MS data were extracted using MassHunter software version B.07 SP1 (Agilent Technologies) and further analysis was conducted using a custom designed in-house built software MSHub. (172)

QC procedures were carried out as described in chapter 3 of this thesis, at all steps in the sampling and analysis process. In brief, TD tubes were cleaned before use with a TC20 conditioning station (Markes Ltd) following manufacturer instructions (40 minutes at 330°C). A sample of 30 cleaned TD tubes, picked randomly from the lab stock of TD tubes, were tested each week to ensure efficacy of the cleaning progress, as part of standard lab QC checks. Breath collection and ReCIVA® sampling QC, laboratory instrument QC and breath quality QC were all performed exactly as previously described, in chapter 3 of this thesis; a threshold of acetone levels of >4,000000 peak area (100ppb) was used to determine "adequate breath sample quality" for samples run using the GC-MS. This was the threshold also used for the COBRA study in chapter 4. However, for the MAGIC study, this same previously described methodology was used to develop thresholds for samples run with the PTR-ToF-MS. This created thresholds of acetone >50 ppb, isoprene >2.5 ppb, isoprene >5 ppb for PTR-ToF-MS H_3O^+, NO^+ and O_2^+ ionisations, respectively. TD tubes that did not meet this minimal threshold were assumed to have had inadequate volumes of breath within them and data were therefore excluded, in the same way as previously described for GC-MS samples. Appendix 24 details the development of the breath content QC for PTR-ToF-MS, as this was not described in detail before.

5.4.4 Training

NIHR research nurses/CSOs were trained about the study itself, the consent process and how to set up and use the breath test at designated Breath Test Training Days held at St Mary's hospital London (27/10/16 and 4/11/16) and at Marylebone Health Centre (November 2017). These were taught by investigator GW, and lasted approximately 2 hours. An instruction booklet was provided with each ReCIVA® portable device, with a step-by-step guide of how to set up the equipment. This was also emailed to all nurses/CSOs prior to the start of the study. The ReCIVA® manual was tested and developed using colleagues who had never set up or used the ReCIVA® before using the Think Aloud Protocol. (173) At all times

nurses/CSOs were able to call study investigators via mobile phone if they had any issues/questions.

5.4.5 Study recruitment phases

The recruitment target was 1000 patients over 12 months of active sampling time. This number was decided upon to represent the approximate theoretical size of future studies planned by the Imperial VOC group. To evaluate different methods of recruitment and sampling the study was divided in to two phases.

Phase-1 (6 months) was designed to:

- Assess the feasibility of four different recruitment strategies (*Opportunistic same day*, *Phone, Text messaging (SMS), and Opportunistic pre-booking)*
- Explore the acceptability of the testing process from a medical professional's perspective
- Evaluate a study design that aims for maximal recruitment

Phase-2 was designed to:

- Explore a centralised sampling model for breath testing (hub and spoke model)
- Explore patient acceptability of the breath test.

During Phase-1 (29th November 2016 to 26th May 2017) samples were collected 'on-site' at selected primary care practices. Two portable breath sampling devices were available, allowing simultaneous patient recruitment in two practices. After each two-week period breath sampling equipment was relocated to two new practices.

During Phase-2 (7h November 2017 to 14th June 2018) a 'hub and spoke model' was trialled; seven practices that were part of the Central London Healthcare GP federation recruited concurrently by referring all patients to a single central practice for breath testing (Marylebone Health Centre), regardless of the patients' usual GP practice. Local collection was also continued at three practices for the latter part of Phase-2 recruitment.

5.4.6 Study recruitment strategies

During Phase-1 of the MAGIC breath testing study, four different models (*Opportunistic same day, Phone, Text messaging (SMS), and Opportunistic pre-booking*) were

trialled for identification and recruitment of patients, with the aim of testing feasibility and acceptability of different recruitment strategies. This was intended to inform about potential ways in which to target patients for a breath test in future clinical practice.

Recruitment model 1: Opportunistic (same-day) sampling

This method was used as the predominant recruitment method throughout Phase-1 of the study (29/11/16 to 26/5/17) in 8 practices. Patients were recruited on the basis of an ad-hoc visit to see their GP. GP practices sampled over 1-2 week blocks at a time. Two separate practices were often sampling concurrently. During these sampling weeks, a dedicated NIHR research nurse/CSO was placed in an available clinic room Monday to Friday, with a breath testing kit. During these allocated weeks, GPs identified any patients they were seeing who were potentially eligible for the study and asked them to agree to speak to the CSO straight after their appointment. If patients agreed to hear more about the study, they went to the adjacent room to have the study explained, have eligibility checked, be formally consented, and perform the breath test as detailed above. Posters were given to GPs at the start of each day to remind them about the study (see appendix 18), and prior to the study all participating GP practices had been sent a letter for dissemination to GPs explaining the study (see appendix 19).

Recruitment model 2: Phone recruitment

This method was used in two GP practices during Phase-1 of recruitment. These practices were chosen as they had the staffing and capacity to telephone patients. This model worked by GPs reviewing their practice list and highlighting patients who may be potentially eligible for the study. Receptionists at the practice then telephoned them to ask if they would be willing to come in to the practice to hear about the breath testing study, with a view to participate. The list of potentially eligible patients was made by the GP practices by screening patient computer records for symptoms or for use of key medications such as proton pump inhibitors (PPIs) or laxatives. Receptionists at the GP surgery rang patients and booked them into dedicated breath-testing clinics. On arrival participants were then checked for eligibility, given information, consented and tested in the same way as described above, by CSOs/research nurses. This differed to the “Opportunistic same day sampling” model

because large numbers of patients (up to 12 in a morning) could be booked into dedicated breath testing clinic sessions.

Recruitment model 3: Text (SMS) recruitment

This model worked in a very similar way to Phone recruitment in that potentially eligible patients were identified by their GP by reviewing the practice patient lists. However, instead of telephoning, patients were sent a text message asking them to reply "yes" if they would be interested in having a call-back regarding the study. 8 practices in Phase-1 one used this as a predominant mode of recruitment, as well as all of the practices in Phase-2. The list of potentially eligible patients was made by the GP practices by screening patient computer records for symptoms or for use of key medications such as PPIs or laxatives. See appendix 25 for a detailed description and example of a search.

Text wording was as follows (one practice used personalised text messaging as this was their usual method of text communication):

"Are you available to donate your breath for cancer research? (..Name..) Surgery are asking our patients to help develop a new breath-testing device. A sample of breath will be collected at the practice during a 15-minute visit with a researcher. You will be helping to develop a new tool that could potentially be used for early cancer diagnosis in the future. Are you interested in hearing more? Text YES for a call back"

We (the central research team) were then able to call by telephone those who replied "yes", to book them in for an appointment. Because it was the trained research team calling patients as opposed to receptionists (as it was in Phone recruitment), this method differed in that patients could ask questions and receive study information on the phone prior to agreeing to come for the full consultation. It also meant that the use of this method was not limited by GP practice staffing (as we were not relying on GP receptionists to spend time calling patients). All surgeries already routinely used texting to contact their patients. Of note, only patients who had previously agreed to receive texts from the surgery were texted for this study (this information had previously been stored in the patient records by the practice).

Recruitment model 4: Opportunistic pre-booking

This described when a patient was seen on the day by their GP and agreed to participate in the study (as per Opportunistic same-day recruitment explained above), but did not want/have time to participate on the day. These patients were then pre-booked to attend for the breath test on a later date in an allocated breath test clinic. This model was used in 2 practices during Phase-1.

5.4.7 Sampling models

Two different sampling models were trialled during the study period, the first in Phase-1 of study recruitment and the second during Phase-2.

Sampling model 1: local practice model. For Phase-1 of the study, two breath testing kits were stationed in two separate GP practices (sixteen practices participated in total), for two weeks at a time. Research nurses/CSOs were on a rota to cover Monday to Friday at GP practices, and used one of the four different recruitment strategies as explained above. Patients were therefore sampled locally at their own practice regardless of how they were recruited. 20 different CSOs/nurses sampled patients over this time.

This local sampling model was used again later during Phase-2 of recruitment at the end of the study because further recruitment was required between 19/1/18 to 14/6/18. This occurred in three further London GP practices (no overlap with practices from Phase-1).

19 practices in total therefore participated in this local model of sampling.

The local sampling model manifested in two ways.

1. Drop-in testing. With purely Opportunistic (same day) sampling, recruitment rates were dependent on GPs referring patients. They were sampled as and when they came, in an unpredictable fashion.

2. Breath testing clinics. In practices using Phone, Text messaging or Opportunistic pre-booking recruitment, patients were pre-booked to attend for their breath test. This allowed the creation of dedicated breath testing clinics. Up to 12 patients could be booked in per half day using this method.

Sampling model 2: Sampling in a hub and spoke model

This model aimed to test feasibility and acceptability of a centralised strategy of breath testing, during the first part of Phase-2 of recruitment (7/11/17 to 18/1/18). Seven practices as part of a London CCG all agreed to recruit for the study simultaneously over the study period. Marylebone Health Centre was allocated as the central "hub" practice, where all patients received the breath test (regardless of their local referring practice). This enabled only two CSOs to be the main breath testers, rather than having to train multiple staff, as patients from all practices were booked into the same dedicated breath testing clinics at one site only. This also meant that all practices could recruit simultaneously over the entire Phase-2 study period rather than having two-week allocated slots, as had been the case during Phase-1.

5.4.8 Assessment of feasibility and acceptability of breath testing

The feasibility of breath testing was assessed using a mixed methods approach. Qualitative assessment strategies were used throughout Phase-1 so that challenges could be identified and overcome in real time, in order to arrive finally at a streamlined and effective breath testing process. Phase-2 of the study then assessed feasibility and patient acceptability of the optimised breath testing process and of the hub-and-spoke model of sampling, using a patient questionnaire (appendix 22).

Phase-1:

- *Field notes* were used to document weekly events and to inform *Plan-Do-Study-Act (PDSA) cycles*. 'Plan' involved creation of a weekly recruitment strategy accounting for surgery-specific considerations e.g. half-days and room availability. 'Do' consisted of sampling. Investigators (GW and the lead CSO) had daily contact with CSOs and recorded verbal feedback of any recruitment, sampling or logistics problems and solutions. 'Study' was weekly review of this process. Finally, 'Act' was achieved by planning with CSOs how to overcome barriers for the subsequent week. The effects of this adaptable study design were part of the study findings.
- A *teleconference* and subsequent *focus group* were used to explore feasibility and acceptability of the testing process, from the viewpoint of the CSOs. This occurred at one month and 6 months respectively, during Phase-1 of the study. The teleconference was

an unstructured CSO-led conversation and feedback session (2 males, 4 females and GW). The focus group (1 male, 11 females and GW) consisted of a brief presentation summarising study progress, then a CSO-led discussion regarding perceived feasibility, acceptability, challenges and mitigation strategies. The focus group was video recorded and later transcribed.

- *Thematic analysis* was performed on all the above transcripts, by coding transcripts to identify main themes. (263) Representative quotes were picked out to illustrate the themes identified.
- *Questionnaires* (anonymous) were given to participating GPs. Likert style questions focused on their opinions around study design and logistics, with open questions regarding the remit of breath testing in primary care (see appendix 23).

Phase-2:

- *Patient acceptability questionnaires* were used to explore opinions about the process, equipment and concept of the breath test (see appendix 22). The design was influenced by other established questionnaires, using Likert scales. (232, 264)

5.5 **Results**

Four main themes emerged from the thematic analysis of the Phase-1 field notes, teleconference and focus group regarding CSO perceptions of feasibility of the study and acceptability of the breath test:

(i) Theme 1: Patient recruitment
(ii) Theme 2: Breath collection
(iii) Theme 3: Breath storage/transport
(iv) Theme 4: Laboratory analysis.

Having introduced these key themes, all results, both for Phase-1 and Phase-2, will be presented under these 4 headings for clarity and convenience, and to avoid any cross-over/duplication of findings. Patient acceptability will be presented under the breath collection theme.

5.5.1 Theme 1: Recruitment

Recruitment was successful, reaching 1002 patients after 192 (of 260 allocated) sampling days. However, 4 patients were later excluded (for being aged >90 years), leaving 998 in the analysis. Patient demographics and reported symptoms are presented in Table 5-1 and range of ages in Figure 5-1. The majority (of those with symptoms recorded) had upper gastrointestinal symptoms (89%) over lower gastrointestinal symptoms (66%), but there was a substantial crossover. This was firstly because the majority of patients reported multiple symptoms (only 165 (18%) complained of only one symptom), but also because the NICE guidelines on gastrointestinal cancers have some cross-over between upper gastrointestinal and lower gastrointestinal symptoms e.g. anaemia, weight loss and abdominal pain. (20, 262) Appendix 17 summarises the symptoms qualifying as "urgent" as per the NICE guidelines on gastrointestinal cancer referral (updated in 2017). Interestingly, 17% and 31% of patients reported gastrointestinal symptoms suggesting need for an urgent upper or lower gastrointestinal referral, respectively. If you take only the patients reporting lower gastrointestinal symptoms, the number is even more striking, where 48% of them warranted urgent lower gastrointestinal referral. A significant level of worry of attending patients was hinted at by the clinical data questionnaires completed by 365 eligible subjects during Phase-2: 10% of patients consulted a GP within 1 week of developing their most troubling symptom, with 28% consulting within 8 weeks. 40% of patients were either "*extremely worried*" or "*quite a bit worried*" about their symptom(s). See Figure 5-2 and Figure 5-3.

Table 5-1 Demographics of 998 eligible patients, and characteristics of reported symptoms

	All patients N=998	Phase-1 N=633	Phase-2 N=365
Age, years (range)	59.7 (18-90)	59.3 (18-90)	58.8 (18-90)
Sex			
Male	409 (41%)	244 (39%)	165 (45%)
Female	578 (58%)	385 (61%)	193 (53%)
Unrecorded	11 (1%)	4 (1%)	7 (2%)
Race			
Caucasian	599 (60%)	335 (53%)	264 (72%)
Asian/Asian British	189 (19%)	161 (25%)	28 (8%)
Black/African/Caribbean/Black British	100 (10%)	73 (12%)	27 (7%)
Arab	30 (3%)	17 (3%)	13 (4%)
Other	60 (6%)	31 (5%)	29 (8%)
Unrecorded	20 (2%)	16 (3%)	4 (1%)
Current Smoker	120 (12%)	72 (11%)	48 (13%)
Oral intake <5hours	798 (80%)	458 (72%)	340 (93%)
Duration of main symptom(s)			
Today	234 (25%)	139 (22%)	95 (26%)
Recently (within 8 weeks)	351 (38%)	241 (38%)	110 (30%)
Chronic	172 (19%)	112 (18%)	60 (16%)
Unrecorded	241 (26%)	141 (22%)	100 (28%)
Patients reporting	**N=921[1]**	**N=586**	**N=335**
≥1 UGI symptom	822 (89%)	533 (91%)	289 (86%)
≥1 LGI symptom	608 (66%)	397 (68%)	211 (63%)
Single symptom reported	165 (18%)	110(19%)	55(16%)
UGI symptoms(s) warranting urgent referral[2]	152 (17%)	98 (17%)	54 (16%)
UGI symptom(s) warranting non-urgent referral[2]	306 (33%)	199 (34%)	107 (32%)
LGI symptoms(s) warranting urgent referral[2]	289 (31%)	178 (30%)	111 (33%)

[1]Symptoms were unrecorded in 77 patients, however for 44 of these patients the 'duration of symptoms' was recorded. [2]UGI, upper gastrointestinal. LGI, lower gastrointestinal. Symptom(s) warranting urgent direct access endoscopic or radiological referral or urgent 2WW referral as per NICE guidelines for UGI (including pancreatic cancer) and LGI cancer. (20, 262)

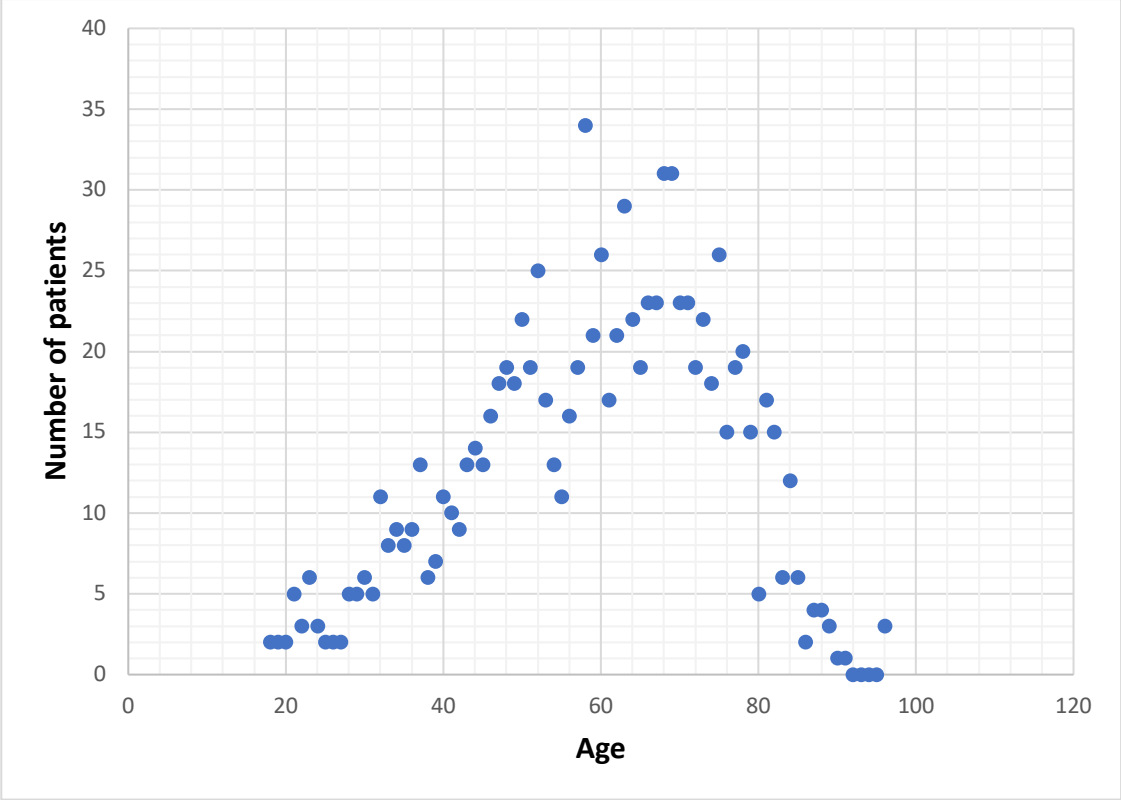


Figure 5-1 Age distribution of all patients with age data (n=980) in the MAGIC study

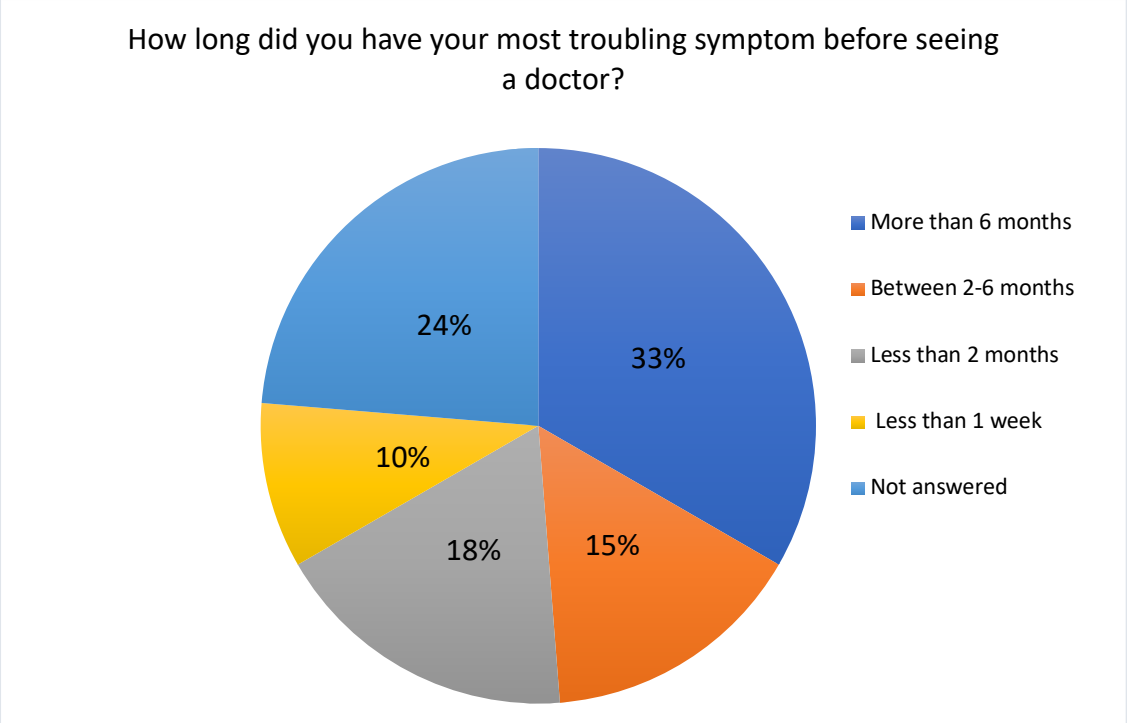


Figure 5-2 Duration of symptoms for 365 patients in Phase-2 of the MAGIC study

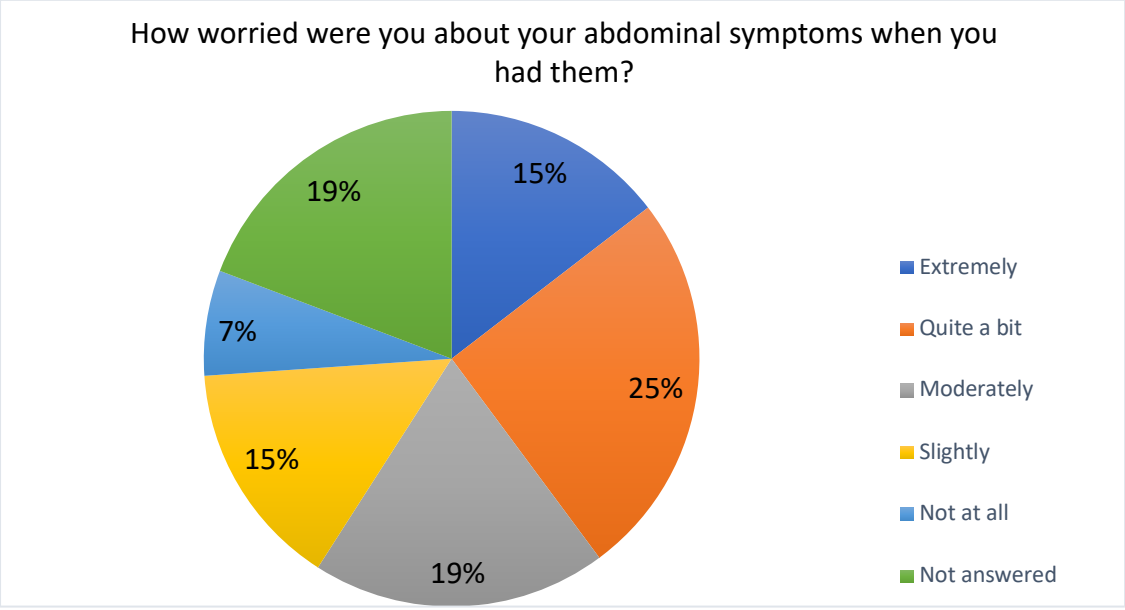


Figure 5-3 Level of worry for 365 patients in Phase-2 of the MAGIC study

Results of different recruitment methods for effective patient engagement

Recruitment was feasible and effective with all four trialled recruitment methods (Opportunistic same day, Phone, Text messaging (SMS), and Opportunistic pre-booking), and using two sampling models (local sampling Phase-1 vs. hub and spoke model Phase-2), see Table 5-2 and Table 5-3. The recruitment method responsible for the most patients was the Text messaging (SMS) method (301 of the 633 patients in Phase-1). This study was not designed to directly compare recruitment methods, but this is interesting because Text messaging (SMS) had the lowest percentage recruitment success rate (11% of all invitees attended), yet this method alone accounted for 48% of total Phase-1 recruitment due to the sheer volume of patients contacted (n=2653 in Phase-1 alone), see Figure 5-4. The implications of this are discussed later. The percentage of patients who completed the breath test after agreeing to be tested ranged from 84% (Phone) to 100% (Opportunistic same-day) depending of the method of initial engagement. Where patients either opted or were required to pre-book a breath test, test completion rates tended to be lower reflecting a 'dropout' rate of between 15% to 18%.

Results of different sampling models for effective patient engagement

In Phase-1, using local sampling at individual GP practices, 633 eligible patients were recruited during 119 sampling days (average 5.3 patients per day) whilst in Phase-2 365 eligible patients were recruited during 73 sampling days (average 5.0 patients per day). Using the 'hub and spoke' model recruitment alone (Phase-2) averaged 4.3 patients per day. During Phase-2 patient recruitment using the 'single practice' model was maintained at 5.3 patients per day. See Table 5-3. When normalised to number of GP practices contributing to patient recruitment for each recruitment method within both Phase-1 and Phase-2, the average number of patients accrued per centre per day was higher for the 'hub and spoke' compared to 'local sampling' method (0.61 vs. 0.28), see Table 5-3.

Table 5-2 Patient enrolment by recruitment method for Phase-1 MAGIC study (633 eligible patients)

	Opportunistic same-day	Phone[1]	Text (SMS) messaging	Opportunistic pre-booking
Number of GP practices Phase-1 (Total = 16)[2]	8	2	8	2
Total number of sampling days that the given method was used[3]	68	15.5	81	15.5
Eligible patients telephoned/sent SMS message[1]	-	114	2653	-
Patients booking an appointment	-	68	345	81
Patients attending booked appointments[4]	206 (100%)	57 (84%)	301 (87%)	69 (85%)
Patient recruitment per sampling day (mean)	3.0	3.7	3.7	4.5
Patients recruitment per practice per day (mean)	0.37	1.84	0.46	2.25

[1]Unanswered/wrong number calls are unrecorded. [2]A total to 16 practices contributed to recruitment during Phase-1, four practices use a combination of face-to-face enrolment and either telephoning or SMS messaging, hence they are counted twice for the purposes of this table [3]Recruitment could occur at two GP practices at any one time, hence, for the purposes of this table only, sampling days could be counted twice, hence total is >192days. [4]Values in parenthesis represent percentage of patients agreeing to breath testing who actually completed the test.

Table 5-3 Patient sampling models used throughout the MAGIC study

	Single centre model	Hub and spoke model
Number of GP practices	19	7
Total number of sampling days	168	24
Total recruitment	895	103
Patient recruitment per day (Mean)	5.3	4.3
Patient recruitment per practice per day (Mean)	0.28	0.61

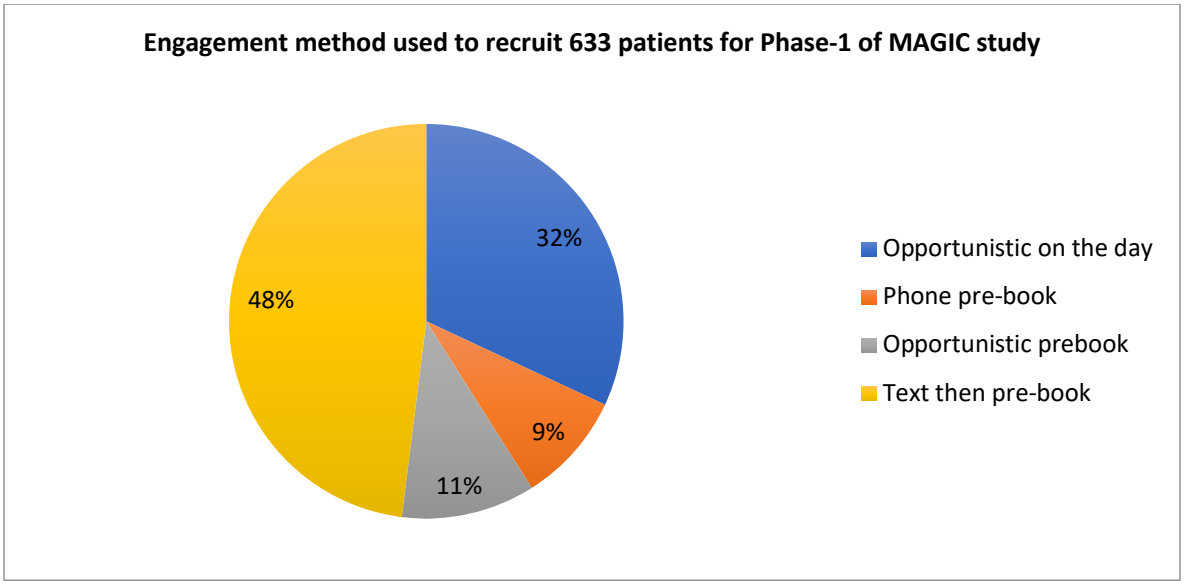


Figure 5-4 Patient recruitment percentage by the method used to engage them, Phase-1

Feasibility of recruitment for a breath testing study

Patient recruitment rate was initially low (Figure 5-5), mainly due to inconsistent referral of patients, inefficient use of CSO time, technical problems and mismatch of CSO and GP schedules, see Table 5-4 for full details of the challenges encountered and subsequent changes that were made. This was followed by a marked acceleration in recruitment during months 2-5 of the study, see Figure 5-5. This was likely partly due to improved CSO familiarity with equipment and study procedures over time, but was also likely due to the dynamic and adaptable study design, driven by weekly PDSA cycles, allowing early recognition of problems and development of solutions. One example of a reactive site-specific alteration was the training of local practice staff to breath-test in practices that had many evening/out of hours clinics. This meant that patients were no longer missed, despite CSOs being unable to sample at these times. After recruitment acceleration, the rate then stabilised and was maintained, even after the sampling model was changed to the 'hub and spoke' model at the start of Phase-2 (Figure 5-5). This finding revealed that testing patients at a centralised site did not negatively affect recruitment. It also indicated that a dynamic and responsive study design may be an effective strategy for primary care studies like this, as recruitment was maintained despite using 26 practices with different environments and clinic schedules.

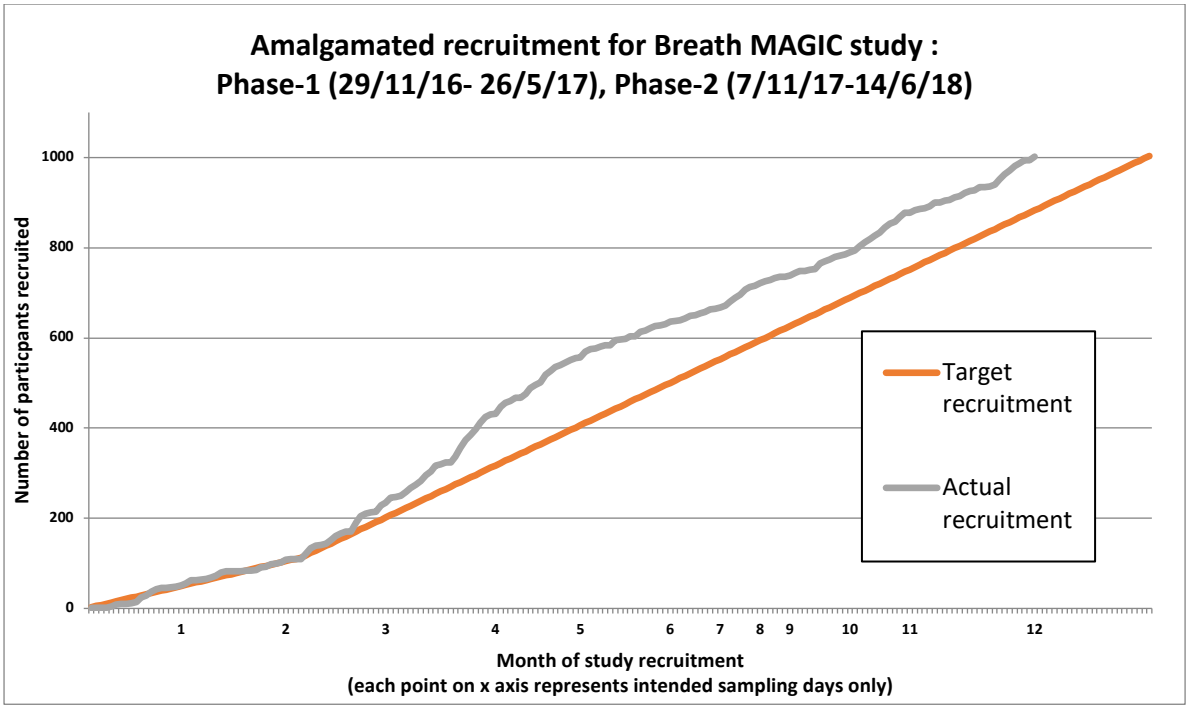


Figure 5-5 Recruitment for the MAGIC breath testing study

Table 5-4 Summary of barriers to recruitment and subsequent implementations, during Phase-1 of the MAGIC breath testing study.

Data in this table is from the thematic analysis of PDSA cycles from field notes, a teleconference and a focus group.

Recruitment Barrier	Cause of this problem (CSO view point)	Implementation made	Effect of implementation	Resolved?
1. Slow referral of patients, leading to slow recruitment and inefficient use of time for CSOs (6 of the first 10 days of sampling had 0 or 1 referrals per day only).	-Busy GP practices: GPs didn't have time. - GPs varied in their interest levels in research. -Locums were less likely to refer. -Some had not heard about the study *-"Nurses and health care assistants would refer patients but GPs had to be prompted every session."* *-"6 patients were found in the system to have GI symptoms, but none were referred. When asked directly, GPs and nurses said they hadn't remembered."*	-Emphasis on daily interaction of CSOs with GPs, reminding them of study and inclusion criteria. - New GP poster and information leaflet was made.	-Improved experience of CSOs meant they engaged more with GPs. - Poster and leaflet helped guide discussion and acted as a reminder for GPs.	Partially… - Still a marked variation between engagement levels of different individuals, often heightened by very busy practices.
2. Persistence of above problem, Opportunistic sampling therefore inconsistent between practices.	- As above, problem not fully solved.	-Addition of Phone and Text bookings as well as Opportunistic pre-booking.	-Tailored to practice resources; played to strengths of each practice. -Engaged receptionists and HCAs in the recruitment	-Recruitment increased exponentially. -Target reached 6 weeks ahead of schedule.

		-Inclusion of referrals from other healthcare staff e.g. specialist/practice nurses and HCAs doing clinics alongside GPs.	task- drastically increased numbers. - Efficient use of CSO time with dedicated breath clinics. -However relied on staff time to pre-identify and call/text patients.	
3.Smaller practices had lower recruitment.	-Some small practices had fewer GPs and fewer sessions. This meant that on days that there were baby clinics/other specialist clinics, no sampling could occur.	-Combine practices that are close together, to boost recruitment and make CSOs time more efficient.	-We combined Carepoint and Acre surgery in Northwood, so that patients could be referred from either practice to see the same nurse. This helped recruitment.	-Good solution for these practices, but not possible in every location.
4.Labour intensive sampling: requires a dedicated CSO present all day for sampling. Costly from central research team staff perspective.	-The study design relied on central research team doing all the recruitment and breath testing.	-We trained practice staff to deliver the breath test instead of central study team. Health care assistants, nurse practitioners, local research nurses and a medical student (after GCP training) were trained in 4 practices, which covered the sampling over 10 study weeks.	-This was a huge success. It saved a lot of central study team time/resources. Local teams found it easy to recruit as they knew their colleagues and often knew the patients, familiar with the computer systems and had their own clinic rooms. They also recruited during evening clinics as this was their expected working hours at the GP practice.	Yes – This also showed how a breath test could be used in future by multiple different staff members. -There was also the added bonus of great collaboration with sometimes new GP practices with great engagement from many different staff members.
5.Study timings of 9-5pm did not always match GP clinic hours.	-Some GPs have evening clinics and have admin time in the afternoons- not good for patient sampling.	-CSO staff hours were adjusted where possible. -Addition of "Opportunistic pre-booking" to arrange for patients to come back within hours. -Local practice staff were trained (see above) which meant they could sample in late clinics if this was their usual working pattern.	-This strategy allowed sampling outside our planned sampling hours. -Other out of hours patients could also still be recruited at a later date using Opportunistic pre-booking.	-Yes- recruitment was excellent during the weeks where we used local staff, and it had other positive effects such as cost saving and increased engagement with local practices and different allied health professionals.
6.Equipment shortages and delays.	-Mask supply ran out at one point in the study. -GC-MS malfunctioned at one point, leading to backlog of TD tubes and therefore a supply issue.	-For problematic weeks, sampling was limited to 20 patients per week. This affected 8 of the sampling weeks (6 practices affected). - Samples were processed on the PTR-MS where possible, allowing clearance of back log.	-20 per week cap was not a problem for most GP practices as long as they knew in advance. -Only one GP practice was negatively affected where 3 days of sampling had to be cancelled. -3 weeks of recruitment was significantly affected by this issue because phone bookings were not made in practices where this would have been the recruitment method.	-Resolved at the time with back-up alternative instrument. -Instrument malfunctions are unavoidable but this could also be mitigated by a larger back-up supply of tubes. - Masks will now be ordered a year in advance for next time.
7. Problems with rooms/space.	-Room availability- "*There was a pressure on rooms today, and I had to move rooms mid-course during recruitment and find a suitable computer to work from. The recruitment was fairly low as a result".* -Upstairs rooms –*"being based upstairs as opposed to downstairs where most patients were being seen impacted on recruitment".*	-Requesting downstairs rooms where possible/rooms near where the patients were being seen.	-Not clear what the effect of the intervention was, as room availability could often not be controlled, but anecdotally the nurses felt that more patients came when the sampling room was easy/accessible.	-Partially resolved by enquiring and requesting certain rooms, but room availability was often largely not within our control.

Feasibility of different methods of recruitment

Although all methods of recruitment were deemed feasible by CSOs, they also commented on the advantages and challenges of each, see Table 5-5.

Table 5-5 Feedback regarding recruitment strategies from focus group and field notes

Opportunistic same-day / Opportunistic pre-booking	
Advantages:	• Method of recruitment open to all GP practices • Easy to organise as reliant on direct patient interaction (without need for telephoning or SMS messaging) • No requirement for administrative staff • *Opportunistic same-day*: convenient for patients as no separate visit needed • *Opportunistic pre-booking*: allowed patients to be brought back at a time convenient for them
Challenges:	• Reliant on GP engagement: CSOs having to "remind GPs 2-3 times per morning" with some GPs admitting to "forgetting to send in patients" • Slower recruitment in smaller (less busy) practices • Inefficient: CSOs present all day for a mean yield of approximately 3 patients
Telephoning	
Advantages:	• Method of recruitment open to all GP practices • Appointments available for up to 12 patients per half day 'breath clinic'
Challenges:	• Requires support of administrative staff to contact patients • Administrative staff only able to give general information about study when calling patients • Cost of telephoning (including staff time)
Test (SMS) messaging	
Advantages:	• More efficient than telephone booking as patients who were telephoned had already expressed interest in being involved in breath testing by responding to SMS message. This led to higher booking rates. • Appointments available for up to 12 patients per half day 'breath clinic'. • Patient initial identification not reliant on attendance to GP. • Potentially more convenient for patients.
Challenges:	• Only open to GP practices with ability to send SMS messages to patients. • Proportion of patients who were ultimately tested lower than for other enrolment methods (approximately 10% uptake). • Identification of patients based on database searching has the potential to be less reliable and could vary between different practices. • Requires support of administrative staff to contact patients. • If administrative staff are used to call back patients they are only able to give general information about study when calling patients. • Cost of telephoning (including staff time).

	Local sampling model	**Hub and spoke model**
Advantages:	• Patient convenience in attending own GP practice • Allowed for Opportunistic same-day enrolment	• Broader recruitment cohort • Fewer CSOs required (2 for single hub): reduced CSO training time and potential improved consistency and quality of sampling. • More flexibility for patients wanting to book appointment and more efficient for CSOs to collect samples.
Challenges:	• Narrower recruitment cohort • Larger number of CSOs required, with less efficient use of their time	• Some patients may either not wish to or be able to travel to the central hub for testing. • Allocation of appointments between multiple practices, requirement for central booking system.

Different recruitment methods were assessed with regards to populations reached, to ensure consistent representation of all invited patients without influence of the method itself. Age, gender and ethnicity were explored for each recruitment method.

Opportunistically recruited patients tended to be younger, mean age 55.6, whereas patients who agreed to return for a dedicated breath testing clinic appointment tended to be older. However, the mean was still below retirement age for all, see Figure 5-6. Text responders were the oldest age group, showing that any preconceptions about use of technology in older age groups were unfounded.

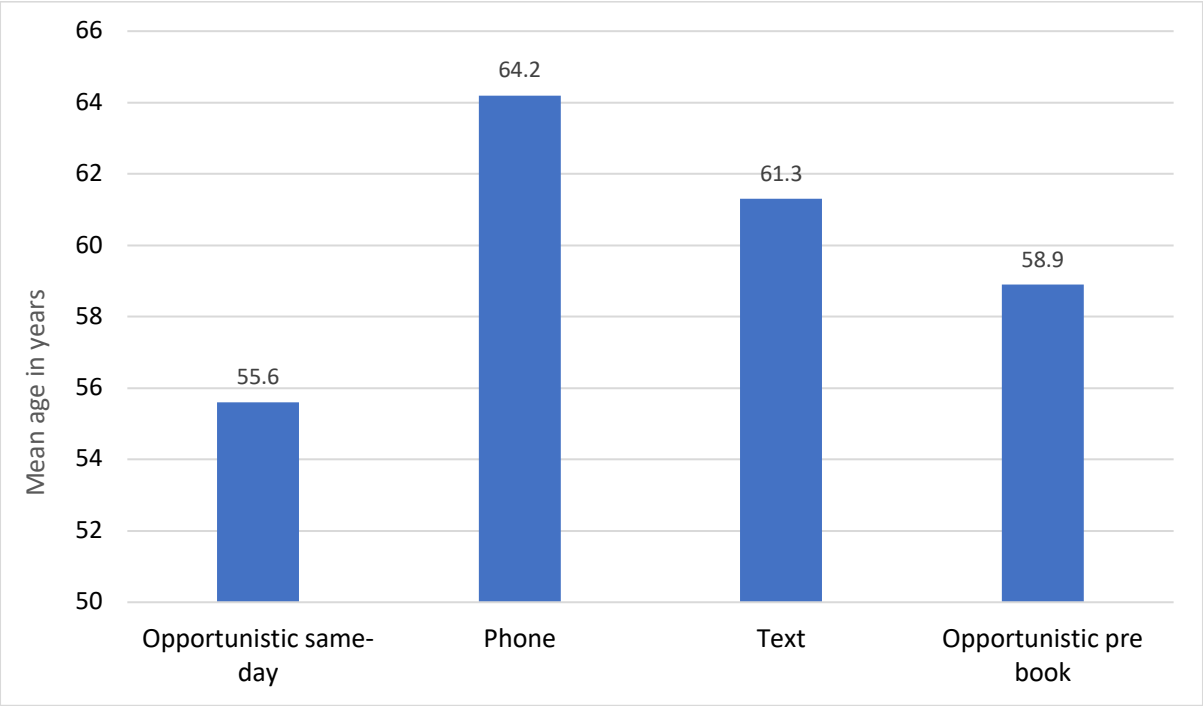


Figure 5-6 Influence of age on response to different recruitment strategies (Phase-1)

With regards to gender, there were more females than males participating in the study, and this was represented by all the recruitment methods, see Figure 5-7. The comparative percentage of males was highest in the Text (SMS) messaging and Phone group, indicating this may generally be a preferred method by males. However, this cannot be fully interpreted as it is unknown how many males and females were approached for the study.

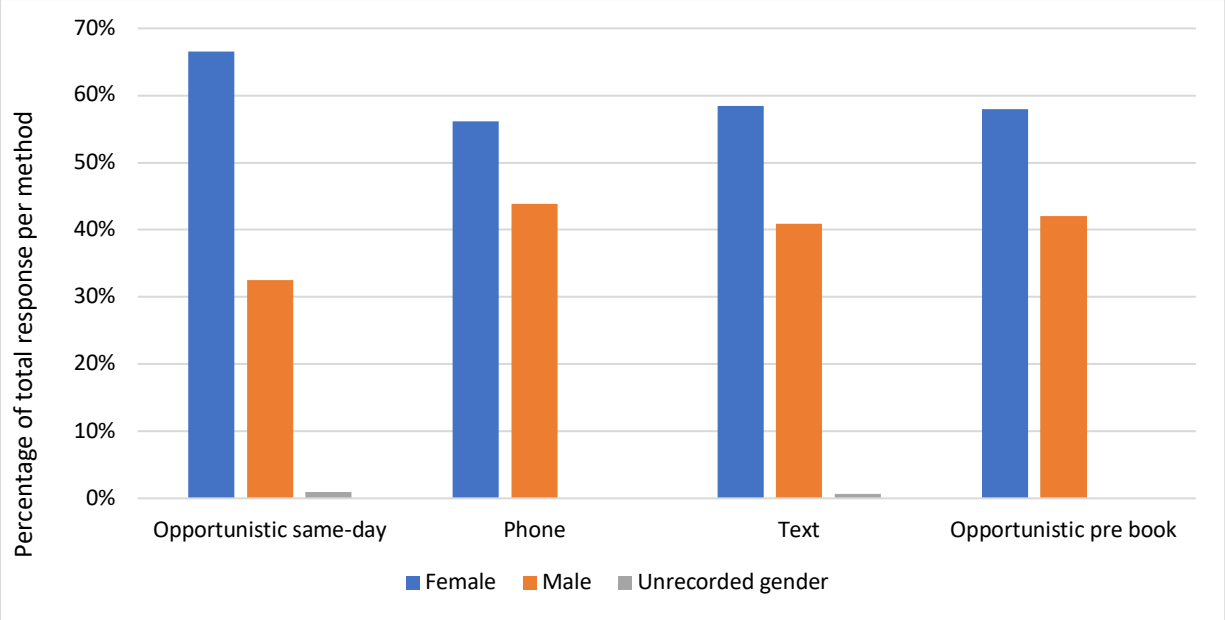


Figure 5-7 Influence of gender on response to different recruitment strategies (Phase-1)

With regards to ethnicity, Caucasian patients made up 60% total recruitment for the MAGIC breath testing study, and this was approximately represented by all the recruitment methods, except for Phone recruitment where there was a higher percentage of Asian/Asian British attendees, see Figure 5-8. However, this cannot be fully interpreted as the Phone recruitment method was used in two practices only, Aksyr and Oxgate practices in Brent, near Wembley (see appendix 16). These were also the same two practices that used the Opportunistic pre-booking recruitment method. Response rates would have been dependent on the demographics of the area, where the ethnicities of all patients approached is unknown.

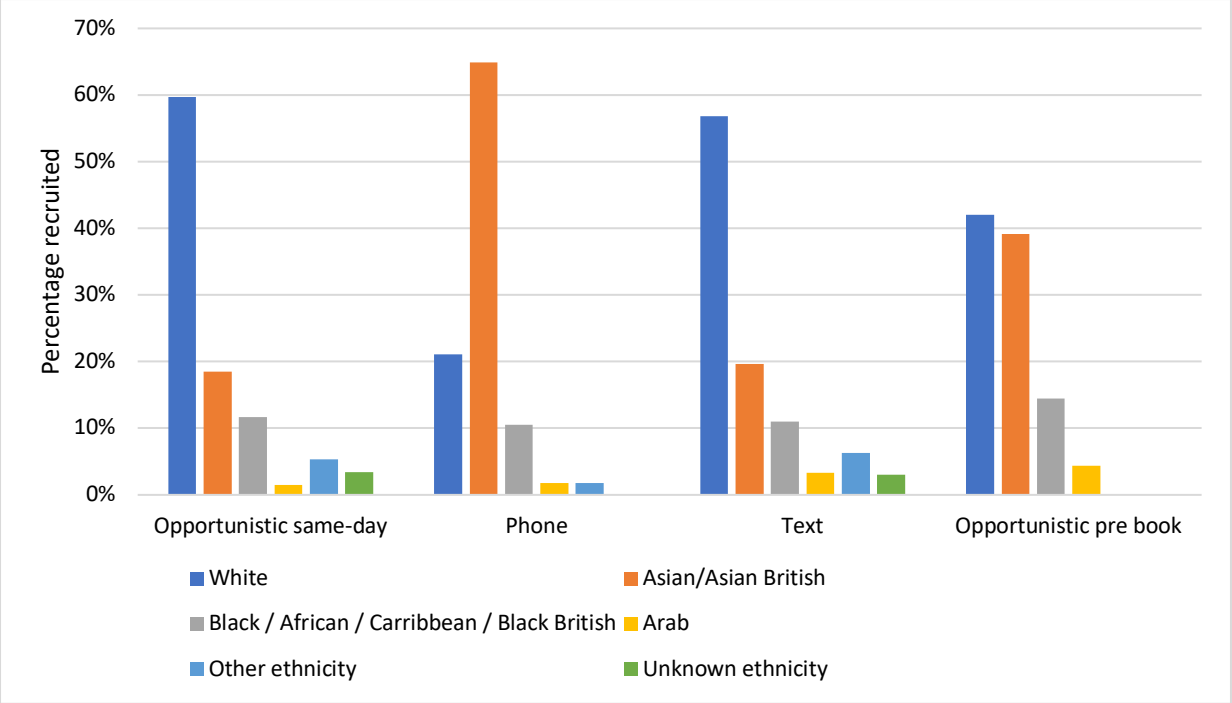


Figure 5-8 Influence of ethnicity on response to different recruitment strategies (Phase-1)

5.5.2 Theme 2: Sample Collection

Twenty-five healthcare professionals were successfully trained to sample breath, showing feasibility of this task for a wide range of operators. These individuals included CSOs, research nurses, GP practice nurses and a medical student (good clinical practice accredited). Most of the information about the collection part of the breath testing process came from the thematic analysis of the teleconference of CSOs, the field notes and the CSO focus group. This was the richest data source as the CSOs were the ones performing the sampling. There were 20 CSOs in total over the first 6 months (12 attended the focus group). In general CSOs found it easy to sample patients, but highlighted some key limiting factors, most of which were equipment related, see Table 5-6 .

Table 5-6 Summary of themes regarding limiting factors of the breath sampling process.

Data in this table is from field notes, a teleconference, a focus group for CSOs, and TD tube processing data.

Theme	**Examples of representative comments**
Patient based limiting factors	"*Small print of the information sheet was an issue for one patient"* *"One patient had a bad cold and felt like they couldn't exhale properly into the mask"* *"One patient was anxious as said she was really claustrophobic, but she managed to do it in the end without a problem"* *" There were really no issues with patients, all were happy to help"*
Equipment (computer) based limiting factors	*"Computer very haphazard. Flowometer not working. About half of the patients were timed samples in the end. Cutting out (going blue) which responded to ctrl alt delete."* *"Flowometer responding to turning off and on but not always. I think it is a mask connection problem,"* *"Large number of timed samples because of flowometer issues"* *"Screen went blue"* *"Screen froze and wouldn't respond"* *"The machine displayed an error message and the breath monitor did not increase. Solved when system was rebooted but patient had gone by then."* *"Computer fault again, not recognising mask and not reading, had to do timed sample. This corrected itself the next day."*
Equipment (ReCIVA® device) based limiting factors	CSOs felt that the equipment was "*fiddly*" but "*ok once they got the hang of it*", particularly the spanners used to tighten screws. They commented on the time it took to set up the equipment at the start of the day and found it easier if there was a desk/workspace to lay out equipment and documents. When commenting on their perceptions of patient acceptability, CSOs said that they observed that patients didn't always breathe "*normally*" when wearing the masks, and some held their breath. This was not reported by patients in the acceptability questionnaires.
Equipment (TD tube) based limiting factors	1 tube in Phase-1 arrived with black soot (sorbent) coming out of one end. This was sent to the company for repacking. Breath data was discarded.
Human errors in sampling	Early in Phase-1 of the study 13 tubes in a batch arrived with no caps on, meaning that the samples would have been very contaminated. This was solved by contacting the CSO individually, (who had mistakenly forgotten to do this). This was an easily solved problem that did not recur. Breath data was discarded. Tubes occasionally arrived with loose caps on. However, these were likely tight enough to have held in the sample, but were easily removed by hand. This was solved by sending a reminder to CSOs and including this as a point in all subsequent training. Breath data was not necessarily discarded, but quality was checked as per all samples. Some tubes were overly tightened with the spanners, which could potentially damage the tubes. This was solved by supplying handheld spanners only, rather than conventional long spanners.

Breath sampling and quality control

Although there were minimal patient-related limitations, Table 5-6 describes the numerous technical issues reported with sampling equipment. Problems were frequently solved by restarting or updating the computer software for the ReCIVA® device. One of the ReCIVA® devices was sent back to the manufacturer for repair. When such measures failed, CSOs resorted to collecting breath as 'timed samples' where patients were asked to breathe into the ReCIVA® for five minutes without using the device's software. This meant that the volume and flow rate of breath sampling was uncontrolled. CSOs did this because they felt they could not ask the patient to wait too long, and patients had often already started sampling when the error occurred. 'Timed samples' accounted for 87 (13.7%) of the 633 eligible samples collected during Phase-1 and occurred primarily at the start of study when CSOs lacked experience using the ReCIVA®. In comparison, during Phase-2 of the study, when study logistics and methodology had been optimised, only 7 (1.9%) 'timed samples' were collected out of a total of 365 eligible samples. During the final six weeks of sampling there were no reported equipment failures.

Acceptability of the breath collection process, from a patient perspective

Acceptability of the breath test to patients was assessed using a questionnaire consisting of ten questions, completed at the time of the breath test (Phase-2 only), see appendix 22. Feedback was overwhelmingly positive. The questionnaires showed that 99% of patients said that the breath test was either 'very easy' or 'easy' to do and 99% of patients felt 'comfortable' or 'very comfortable' wearing the face mask. Similarly, 95% thought that the duration of the test was 'acceptable', and 98% of patients declared that they were 'very satisfied' or 'satisfied' with how the test was explained to them. 99% of patients would do the breath test again if recommended by a doctor, and 98% would 'strongly encourage' or 'encourage' friends and family to complete the test if offered. 77% of free-text comments were to give positive feedback about the test, see Table 5-7.

Table 5-7 Summary of patient acceptability questionnaire responses (n=365, Phase-2 of study)

	Very easy/very comfortable	Easy/ comfortable	Difficult/ uncomfortable	Very difficult/ very uncomfortable	Not applicable
How easy was it to do the breath test? (%)	79	20	1	0	0
How comfortable were you whilst wearing the face mask? (%)	55	44	1	0	0
How did you find the experience of holding the device during the test? (%)	42	55	2	0	1
Would you be comfortable to do the breath test again, if recommended to by a doctor? (%)	64	35	0	0	1
	Took too long	**Acceptable amount of time**	**Too quick**	**-**	**Not applicable**
What did you think about the time it took to give a breath sample? (%)	2	95	3	-	0
	Very satisfied	**Satisfied**	**Dissatisfied**	**Very dissatisfied**	**Not answered**
How satisfied were you with the explanation given for how to do the breath test? (%)	80	18	0	1	1
	Strongly encourage	**Encourage**	**Discourage**	**Strongly discourage**	**Not applicable**
Would you encourage family and friends who were offered a breath test to complete it? (%)	59	39	0	0	2
FREE TEXT COMMENTS REGARDING OVERALL SAMPLING EXPERIENCE					
"Nothing to improve because there is nothing to it. It's nice"					
"I found the breath test to be extremely satisfactory, I am happy to participate in more research"					
"I found it fine as it is. And, it was a rather nice experience. I liked it."					

Other free text comments included thirteen patients who suggested that it would be better to have a hands-free breath sampler, either for '*ease*' of use or because of being '*tired*' by having to hold it up to their face. These comments were offset by other patients saying they liked being "*in full control of the mask*". Three further patients asked on the feedback form whether the disposable mask used was cleaned between patients (it was a single-use mask in sterile packaging), indicating an explanation issue rather than an equipment issue.

Of note, only one (Opportunistic same-day sample method) patient commented on the way they had been recruited: "*convenient that the test was available immediately following my GP appointment*". No one mentioned their experience of the Text (SMS) messaging or Phone recruitment model. Only one patient commented that they had had to travel to a practice that wasn't their usual practice "*my Doctor's surgery is closer to my home*". The lack of any

other feedback about travel implies that the 103 patients recruited via the hub and spoke model did not mind travelling to another GP practice for testing (although this was not specifically asked on the questionnaire).

The breath test was also acceptable to patients with a wide variety of medical problems, including 197 patients with asthma, chronic obstructive pulmonary disease, or other lung disease recorded.

Acceptability of the breath collection process, from a CSO perspective

CSOs felt that the equipment was "*fiddly*" but "*ok once you got the hang of it*", particularly the spanners used to tighten screws. They commented on the time it took to set up the equipment at the start of the day and on the importance of a provided desk/workspace to lay out equipment and documents. The main limitations to acceptability were software/computer malfunctions, as mentioned earlier in Table 5-6. CSOs commented in the focus group that training was vital to performing the breath test because "*preparation is key"*, *"if you prepare before then it runs like clockwork and you can sample patients back to back"*. The troubleshooting manual was "*useful"* but the training allowed "*hands on practice*", "*the fiddliness and multiple steps required made it not very obvious what to do next unless you had had the training*."

When commenting on their perceptions of patient acceptability, CSOs said that they observed that patients didn't always breathe "*normally*" when wearing the masks. This was not reported by patients in the acceptability questionnaires.

Acceptability of the breath collection process, from the GP perspective:

21 GPs answered the GP questionnaire, from 10 of the 26 practices that participated in the study (see appendix 23). GPs said that asking patients to participate in the breath test study (100% of GPs), answering patient questions about the study (95%) and general logistics of the breath testing (95%) was "*easy* "or "*very easy*". Perceived barriers to participation were "time constraints" (both clinical staff and patient's) and the fact that patients would not get any results from their breath test during the study. All GP respondents reported that they had "*no concerns about the study*" from their patients. GPs felt a breath test would be best placed as a point of care test with instant results ("*very useful*" 86%, "*useful*" 9%), and 48% felt it would be useful even if results were available electronically at a later date. There were varied

views about which groups of patients would benefit most from a breath test, ranging from "*all age, any group*" to "*low risk cancer patients without red flag symptoms*", to "*at risk groups*". Of the eleven GPs who said "*at risk groups*", six of them qualified this with specific symptoms ("*dyspepsia/weight loss", "chronic reflux", "chronic dyspepsia", "elderly and frail", "elderly with weight loss", "lower abdominal or upper gastrointestinal symptoms"*) and three others gave age cut-offs ("*>30", ">45" and "45-74*"). Only two GPs gave an opinion about cost per test ("£10" and "£14") where the others indicated "*don't know*" or said that the CCG should decide. GPs were not asked about their reasons for this cost opinion, or what it was based upon. From the GPs' perspective it appeared the breath test was feasible, but that its cost and place in a future referral pathway was yet to be determined.

5.5.3 Theme 3: Storage and transport

Feasibility of storage

Storage was not highlighted as a practical or logistical problem during the focus group, in field notes or during the teleconference of CSOs. The TD tubes were stored in small sealed Tupperware ™ boxes at room temperature at the GP practices. Storage time was measured at two points for each TD tubes used for breath sampling; time between conditioning (cleaning) and sampling, and time between sampling and analysis. This was deemed an important feasibility assessment because storage time could impact upon quality of breath samples.

Tube processing data

Pre-sampling: The mean time between conditioning (cleaning) of a TD tube and its use for sampling was 4.6 days, range 0-21 days.

Post sampling: The mean time between sampling and analysis was 2.8 days, range 0.11 to 11 days. In addition, in Phase-1 83 tubes (13%) were stored by freezing in a -80 degrees centigrade freezer on arrival, because of instrument downtime, and analysed up to 14.1 days later (maximum).

The relevance and acceptability of the pre- and post-sampling storage times need to be investigated further before planning of any future studies. However, given what is in the literature, these seem to be very acceptable storage times, with storage conditions consistent with other studies. (144)

Feasibility of transport

Transport of the equipment was highlighted as a problem by one of the CSOs as it was "difficult logistically" because the "bag of all the masks was very bulky and 20 tubes were heavy" (equipment weighed up to 16kg). Couriers were therefore used to do the main equipment moves. After this there were no transport issues highlighted by CSOs.

5.5.4 Theme 4: Breath analysis

Of 998 included breath samples, the GC-MS was used to analyse 318 samples and the PTR-ToF-MS was used to analyse 680 samples; 385 using the H_30^+ ion source, 285 using the NO^+ ion source, and 10 using the O_2^+ ion source. It made no difference in terms of efficiency/instrument time/analysis time as to which ion was used to run a given sample on the PTR-ToF-MS, hence the ionisation mode selection was arbitrary except to measure feasibility. It was feasible to analyse large numbers of TD tubes using these instruments, except where instrument downtime occurred. This affected a period of two weeks, where TD tubes had to be stored in the freezer before being analysed later.

Quality of sampling process

Sampling process QC parameters were assessed in samples analysed by PTR-ToF-MS (n=316) and GC-MS (n=23) in Phase-2 of this study. Twenty-six additional Phase-2 samples analysed by GC-MS could not be assessed for QC because they were lost due to a GC-MS instrumental error.

Three hundred (95%) of the TD tubes analysed by PTR-ToF-MS and 21 (91%) of the tubes that had been successfully analysed by GC-MS were deemed to contain adequate quantities of breath based on defined sampling process QC criteria, see Table 5-8.

Table 5-8 Pass rate of breath sampling process quality control in Phase-2 of the study

Analysis mode	Number of samples run for analysis	Number of samples lost due to instrument/analysis faults	Number of samples successfully analysed	Number that passed QC sampling process criteria
PTR-ToF-MS, H_3O+ ionisation	145	0	145	136
PTR-ToF-MS, NO+ ionisation	161	0	161	156
PTR-ToF-MS, O_2+ ionisation	10	0	10	8
TOTAL PTR-ToF-MS samples	**316**	0	**316**	**300**
TOTAL GC-MS samples	**49**	26	**23**	**21**

Interestingly, of the 16 samples that failed the sampling process QC for PTR-MS, 5 were "timed samples" indicating that in these cases this was likely the cause for the poor-quality breath sample collection. None of GC-MS run tubes were "timed" samples. It is worth noting that no samples from Phase-2 of the study were frozen, regardless of analytical instrument used. A 91-95% pass rate for good sampling quality over the two instruments indicates that the process used for the breath test was generally feasible and acceptable.

5.6 **Discussion and limitations**

The analysis of exhaled breath VOCs could allow development of a new tool for non-invasive detection of a number of diseases including gastrointestinal cancers. (102) A test like this could be offered to patients in primary care presenting with non-specific symptoms that do not meet existing referral guidelines. However, before a large phase-III primary care-based clinical trial can be conducted, the feasibility and acceptability of the breath test in this setting must be understood. This current study was intended to evaluate possible recruitment and engagement strategies for breath testing in primary care. Phase-1 evaluated different engagement methods, and also investigated optimum logistical and implementation strategies. Phase-2 evaluated acceptability of the test from a patient perspective. The acceptability evaluation was performed subsequently, after optimisation of delivery during Phase-1, so as to obtain most representative results.

The adaptable and flexible study design using PDSA cycles allowed for early recognition of problems and timely development of solutions. This meant that changes were made every week regarding optimising recruitment and study set-up. The use of the teleconference and focus group gained further qualitative data from open discussions guided by the key deliverers of the breath test in day-to-day practice. Focus groups are often used in marketing to gain perceptions, opinions, and attitudes towards a new product. (265) Other acceptability studies for gastroenterological interventions (e.g. Cytosponge) have also used focus groups with thematic analysis for data collection, providing a precedent for this study. (266) It is reasonable to infer that the success of the recruitment in Phase-1 of the study was largely due to this flexible and adaptive study design, based on qualitative data, leading to innovations like the 'breath testing clinic' and the training of a wider variety of healthcare professionals. Phase-1 of the study reached its 500 patient target 6 weeks before schedule. A flowchart of how to optimise patient recruitment in primary care studies, taken from lessons learnt during the MAGIC study is detailed in appendix 26. This study has therefore not only shown that recruitment is feasible in primary care, but also provides a framework that is applicable to other primary care studies.

The MAGIC study also demonstrated that sampling in a single GP practice and sampling in a non-local centralised location were both viable and acceptable options to patients and study staff. The centralisation of breath testing reduced staffing and equipment requirements, with no discernible negative impact on patient feedback. Transport and logistics were simpler from a single location, and the bulk collection it allowed led to lowered courier costs. With regards to organisation within primary care services, breath testing as a one-off test could be compared to a simple blood test. Blood tests are often ordered by a GP but the blood is taken in a different (centralised) location. If breath testing were a complete streamlined service encompassing testing, receipt and interpretation of results, and any referrals required to secondary care, we could draw comparisons to other centralised NHS services, such as the diabetes care pathways. Studies have shown that managing diabetes via GP practices with a specialist interest in diabetes (i.e. centralising diabetes care within general practice) results in equal quality but lower cost than referring to secondary care clinics. (267) The Audit Commission review on diabetes services in England and Wales 1998-2001 supported the idea of more specialist care being provided in a primary care setting (for

uncomplicated diabetes). (268) This study's "hub and spoke model" data shows that a similar centralised referral pathway could potentially be used for breath testing.

Four methods of recruitment were evaluated in Phase-1, (Opportunistic same-day, Phone, Text (SMS) messaging, Opportunistic pre-booking), with advantages and disadvantages previously discussed. All were feasible and acceptable. Age, gender and ethnicity was not shown to have a significant influence on uptake via any method. This means that for future studies, any or a combination of all methods could be used to recruit patients, depending on the intended purpose of the test. For example, text (SMS) messaging, and to some degree phoning, has the potential to reach many patients. However, as highlighted by this study, this approach may only result in assessment of 10-50% of the patients, which is comparable to population screening (the gFOBT for CRC screening had a 54% uptake). (53) Alternatively, opportunistic identification of patients could be done by their GPs, which may be more representative of a targeted triage test approach. A triage tool such as this could be used as an adjunct to existing referral pathways, such as the 2WW pathway. Uptake may be high in this group, given that in this study 40% of patients were either "extremely worried" or "quite a bit worried" about their symptoms and the acceptability of the test was extremely high. However, the place that breath testing has in clinical practice and the method used to engage patients is clearly highly dependent on the result of large-scale diagnostic accuracy studies, currently ongoing, as well as the population targeted.

This study, although aimed at feasibility and acceptability, gave an interesting snapshot into the gastrointestinal symptoms of 998 patients in London attending primary care. A strikingly high percentage of patients fulfilled the eligibility criteria for 2WW or urgent endoscopies, where 17% of those with upper gastrointestinal symptoms and 31% of those with lower gastrointestinal symptoms appeared to be eligible. This provided an approximate indication of the burden of referrals and workload on the NHS, where an effective breath test could potentially alleviate this by streamlining referrals, rationalising the service and potentially saving money. However, this referral data must be interpreted with caution as it is not known how many of these patients were in fact referred on the 2WW pathway, as follow-up data was not collected. Also, the sample is not representative of all GP attendees with gastrointestinal symptoms, as approach and decline rate for the study is unknown, and the practice electronic searches were not designed to be exhaustive. Finally, clinical lab data such as haemoglobin levels or platelet levels of patients were not collected, where anaemia

and presence of raised platelet count (in combination with other symptoms) both feature in the NICE guidelines for gastrointestinal cancer referral. (20, 262)

The breath test had very high acceptance rates. The vast majority of patients found the test easy to complete, where patients represented different age groups, and varied in gender, comorbidity and ethnicity. There may, however, be a selection bias because enrolled patients were those who were most likely to attend their GP in the first place and engage with medical research. In addition, more women than men participated in this study, potentially influencing results. This may be representative of attendees to primary care in general; a Dutch study reported that women are 18% more likely than men to consult their GP, after adjusting for gender-specific factors such as maternity. (269, 270) It is not known for this study how general attendance to the GP and study approach-rates were influenced by gender. There was also a gender bias in the focus group used to evaluate the study methods because there was a higher preponderance for female research nurses/CSOs allocated to the MAGIC breath testing study from the NIHR clinical research network for Imperial College London. This could have influenced results.

A further limitation of this study, which could affect feasibility of breath testing in future, was unreliability of breath testing equipment, mainly at the start of the study. If all "timed" samples (performed by CSOs when the software failed) had had to be discarded due to poor quality, this issue alone could have accounted for loss of 13% of Phase-1 samples (decreased to 1.9% of samples in Phase-2). However, during the last six weeks of the study, after optimisation of staff training and streamlining of sampling methodology, there were no further technical failures of breath collection. Analysis of a high volume of breath samples was feasible using the PTR-ToF-MS and GC-MS in our lab, where PTR-ToF-MS showed its value in managing high-throughput studies such as this one. This is supported by other studies. (142) QC procedures were used to ensure instrumental consistency. (102) Ninety five percent of all samples that were analysed passed the threshold for acceptable quantities of the VOCs acetone or isoprene (see chapter 3 of this thesis for full explanation), and were therefore deemed to contain adequate concentrations of human breath and were therefore deemed 'adequate quality'. For implementation of large-scale breath testing, there would need to be standardisation across different laboratories. Equipment checking and function would need to be an important consideration for future use of breath testing in clinical practice. The

development of point of care devices may assist in streamlining the analysis and receipt of test results, and make the process more efficient.

It has been previously reported that both financial pressures and lack of time can be major barriers to the ability to conduct high quality primary care research. (271) GPs and research staff in this study were very supportive of conducting this research, but this may not necessarily be representative of their views for a future breath test in clinical practice, where GPs would hold the responsibility of implementing testing, interpreting results and then acting on the findings. In this study, factors such as the recruitment of patients via the Text (SMS) messaging route and use of a central sampling hub staffed by study staff all helped to reduce the workload of GPs. In this research setting, GP practices also received a financial incentive from the NIHR; they received £20 and £25 per patient for Phase-1 and Phase-2, respectively, for every recruit. This may have encouraged GPs to participate and made GPs feels that they had been recognised at least in part for any addition to their workload caused by the study. These factors may not be the case for a breath test outside of a research setting, which may influence GP acceptability of the test. The fact that patients were not required to follow any specific conditions leading up to the breath test, such as being nil by mouth, may also make this study non-representative of a future breath testing pathway. However, as there are currently no evidence-based guidelines for performing breath sampling in clinical practice, it is not clear what (if any) specific conditions there may be in future.

No previous study in the literature has investigated how breath testing could be successfully integrated into primary care, with engagement from patients and clinical stakeholders. This study was therefore the first of its kind in the literature. Strengths were its two phased design and flexible mixed methods approach, allowing it to adapt recruitment and sampling strategies according to performance and feedback, as the study ran. Limitations were that the demographics of the patients who declined to take part or who did not respond to the invitation were not recorded, therefore barriers to patient participation were not investigated. We hypothesised that time pressures and the fact that there was no direct clinical benefit to participants of the study may have contributed to the decline rate. However, this should be investigated in future studies to determine the degree of uptake of the test and the factors that influence patient uptake. As part of the GP questionnaire, GPs were asked what role the breath test could have in clinical practice, however without knowing the diagnostic accuracy of the breath test this may be an unfair question to ask. GPs were also

not asked about potential challenges to the adoption of the breath test, which could have provided useful insight. This should be investigated in future studies, ideally when the diagnostic performance is known, so that GPs can form a more informed opinion.

5.7 **Conclusion**

This study determined that it was feasible to recruit large numbers of patients with gastrointestinal symptoms for breath testing in primary care. Recruitment was feasible to both staff and patients using Opportunistic same-day recruitment, Phone recruitment, Text (SMS) recruitment and Opportunistic pre-booking recruitment, and sampling was made more efficient by use of dedicated breath testing clinics. A hub-and-spoke model of centralising breath testing was also feasible and acceptable to both staff and patients, and had benefits in terms of logistics and efficiency. Sampling was highly acceptable to patients and clinical staff. Storage of samples was feasible at room temperature, and samples were readily transported to the lab for analysis within an acceptable amount of time.

This study provided encouraging new evidence to support the implementation of wide-scale breath testing in the primary care setting, running in parallel to much-needed diagnostic accuracy and standardisation studies. It also provided a useful practical framework to guide the designing and set-up of future (larger) phase III trials examining the breath test's performance in this well-known challenging research sector.

5.8 **Funding**

The research was supported by awards from Rosetrees and Stoneygate Trusts via Imperial College Charity (grant number 171812) and by infrastructure funding from the NIHR London In Vitro Diagnostic Co-operative programme (grant number P67522).

6 CHAPTER 6: SUMMARY, CONCLUSIONS AND FUTURE WORK

6.1 **Summary**

This body of work started by reviewing the literature surrounding the use of VOCs for detection of colorectal diseases including CRC, identifying that there was strong potential for the use of VOCs in breath as a diagnostic tool, but that there was also a clear need for a standardised approach to VOC analysis and QC of results. In response, the next section of the work of this thesis followed the journey of the creation of a pragmatic and reliable workflow for the sampling, collection and processing of breath samples using TD tubes, including development of a novel threshold-based QC system for GC-MS and later for PTR-ToF-MS (but applicable to any analytical platform). Using this system, the abundance of acetone or isoprene on each TD tube was determined, using simple peak area count measurements. Where the abundance was below that expected from a human breath sample, the tube was deemed unlikely to be a representative breath sample, due to either a problem in the sampling or storage process. The sample was then confidently excluded before detailed analysis took place, saving time and resources, whilst providing a pragmatic easy-to-follow step towards the goal of achieving high quality and reliable breath testing results. This study fills a gap in the literature for testing the quality of breath samples stored on TD tubes and will enhance the field with regards to working towards a standardised approach to research using VOCs in breath. Following on from this, and thereby putting the QC system into practice, came the successful execution of the COBRA study, the largest CRC breath study in the current literature. This showed that breath VOCs can detect CRC with a high sensitivity and specificity of 77% and 87% respectively, using GC-MS. Polyps of all risk categories could also be detected with a sensitivity of 66% and specificity of 54%. Included in the top 25 predictive VOCs for CRC were 15 VOCs with highly plausible mechanisms for a role in CRC, from the ester, alcohol, alkane, sulphur compound, phenol and non-aromatic cyclic hydrocarbon chemical groups, making them excellent candidates for CRC biomarkers. These 15 VOCs together were able to predict the presence of CRC from positive and negative controls using breath with an area under the ROC curve of 0.83. The abundance of each VOC was also closely related to the T stage of the CRC, supporting the idea that greater metabolic changes created by a larger more invasive tumour may be directly reflected in breath. When compared to currently used CRC and adenoma screening and detection tools, namely the FIT test, the breath test performed well for both screening and symptomatic groups, for both CRC and polyp detection, see chapter 4 discussion section 4.5.7. A further consideration, in comparison to the FIT, is that

the breath test may be more acceptable to patients and therefore lead to improved uptake. If this were found to be the case then this may give it a very important place in the arsenal of tools for CRC detection, either instead of or in combination with the FIT. Following on from the COBRA study, the work of this thesis moved on to examine the wider picture of the concept and design of breath testing in primary care, with the ~1000 patient MAGIC study showing feasibility and high acceptability of the breath test from the perspectives of both patients and healthcare staff. This study showed that a single central lab was able to process large numbers of TD tubes and analyse them within the recommended time frame, and that logistics of transport and storage were feasible. The study also provided a useful framework for the recruitment of large numbers of patients from a primary care setting, valuable for future studies.

The three main studies of this thesis therefore followed a journey of development and progress of breath testing for CRC and colorectal polyps, moving through the subject areas of quality of sampling, diagnostic accuracy, then feasibility and acceptability of breath testing and recruitment of large numbers of patients. The encouraging and exciting findings pave the way for a definitive large-scale breath testing study, which could incorporate direct comparison to the FIT and investigation into the best role for breath testing in a real-life clinical pathway. There were, however, limitations to the main findings of this thesis. The effect on VOCs of potential confounders such as age, bowel preparation and environmental factors could not be fully ascertained from the COBRA (or MAGIC) data, and although T stage of the CRC appeared to have a direct (sometimes linear) relationship to the concentration of each discriminative VOC in the COBRA study, it was not possible to tease apart any effect of environment on this. A further limitation was that the 15 top VOCs found to be discriminative may not be an exhaustive list, as the GC-MS analysis was limited by the fact that a single mid-polar column was used, which would not have detected highly polar ions such as short chain fatty acids. There was also no quantification of VOCs as this study was an untargeted analysis, where calibration curves would have been required from the outset in order to quantify VOCs. Given that the identity of discriminative VOCs was not known until the end of the study, this was not possible, and calibration curves cannot be done retrospectively. There were also limitations with the TD tubes and particularly the ReCIVA® breath sampler, explored in detail in the MAGIC study, making it impossible to eradicate some variability in VOC measurement. Sampling inconsistencies, resulting in failure of QC procedures, was the cause of the exclusion

of 227 patient breath samples in the COBRA study. A further limitation was the lack of testing of repeatability of VOC measurements. Future work addressing these short comings could greatly improve the reliability and quality of results, so that fewer data are excluded. COBRA should be considered as the first part of a larger body of work, where future studies are required to strengthen and validate results.

6.2 Future work

Main future avenues of research highlighted by this work include the further improvement of quality and standardisation of breath testing, working towards the eventual achievement of national standards within a widely agreed approach to breath analysis. The Imperial VOC group continues to work on each step of the breath analysis process; preparing and monitoring study equipment, breath sampling, processing samples, storage, and optimising mass spectrometry analysis methods. Of particular importance is finding novel ways to sample breath in order to optimise sample quality and reduce the need to exclude data. New sampling methods, including external and in-house designed prototypes, will help to address many of the limitations of breath sampling that arose using the ReCIVA® and TD tubes in the COBRA and MAGIC studies. The Imperial group will publish a workflow for each step of the breath sampling process using SIFT-MS this year; the paper has been accepted to Nature Protocols.

Other future work avenues focus on the continued pursuit of a definitive CRC breath test, founded from the positive results of the COBRA and MAGIC studies of this thesis. This requires multiple future studies; firstly, the COBRA 2 study. This is planned to be a large multi-centre study consisting of four sections; work package (WP)-1, a 750 patient study as a direct validation study following the results of the COBRA study presented in this thesis, conducted in the same way. COBRA 2 will adopt a more sophisticated sample analysis using two separate GC-TOF-MS replica instruments, allowing analysis of breath samples with both polar and non-polar columns, optimising capture of VOCs from different chemical classes and providing robust structural identification. Chemical validation of all discriminant VOCs will be additionally confirmed by chemical standards. A definitive wider panel of VOCs may lead to higher accuracy of detection of CRC. As a further improvement, the more sophisticated instruments will also allow quantification of discriminatory VOCs, enabling establishment of a test threshold for separating patients with CRC from controls. The result will be the

refinement of the diagnostic model and development of an algorithm for CRC/non-CRC. WP-2, will be a 5700 patient blinded multi-centre clinical study of all patients attending for an urgent 2WW pathway colonoscopy. This will be an external validation study of the CRC-specific VOCs identified in this thesis and in WP-1, using the resultant diagnostic model. WP-3 will involve comparison of diagnostic accuracy with the FIT, thereby helping to establish where a breath test might best fit on a real-life clinical pathway. This will be done by asking all patients from WP-2 to give a stool sample, to be brought with them when they attend for their colonoscopy (the day of the breath test). WP-4 will be the performance of a stakeholder analysis and health economic modelling aiming to investigate barriers for clinical adoption and develop strategies for scaling-up breath testing for clinical practice. This will build on results from the MAGIC study. Further studies are also required to investigate the molecular mechanisms and exhalation kinetics that lead to the production of VOCs in CRC, to provide a plausible scientific backing and add strength to the findings.

Once diagnostic accuracy of the breath test is established from COBRA 2, both alone and in combination with FIT results, one ultimate long-term goal would be to use the test thresholds to develop a sensor to detect only discriminatory VOCs above set concentrations. Sensor technology may prove to be a cheap and highly acceptable way of testing in the community, given the potential for avoidance of any transport or analytical steps, quick processing times and clear yes/no results. A sensor study could then be performed to ensure diagnostic accuracy was maintained, whilst investigating the uptake and deliverability of such a tool. In time, once the diagnostic accuracy, VOC identification and establishment of thresholds of VOC breath testing has been confirmed for other cancer types, a breath test (sensor or otherwise) could be used to diagnose/rule out multiple cancers in one sitting.

6.3 **Conclusion**

In conclusion, this research has identified VOCs that are significantly altered in CRC and to some degree in colorectal polyps, and that breath testing in primary care is both feasible and acceptable. The results give an exciting boost to the field of VOC research and warrant the planning and execution of several further avenues of research as described above. Breath remains a very attractive and promising medium for the detection of CRC, being non-invasive, quick and likely more acceptable to patients for sampling than blood or stool-based tests. This work, and the future work to come, paves the way for the development of an effective tool

for non-invasive testing for a cancer that currently claims the lives of many thousands of patients per year, lives that could have been saved by early detection and treatment.

7 APPENDIX

7.1 **Appendix 1: Modification of QUADAS-2 assessment tool**

		Existing QUADAS-2 assessment tool	**Modified QUADAS-2 assessment tool**	
Risk of bias	Patient selection	Was a consecutive or random sample of patients enrolled?	Were sampled patients representative of the intended population?	**(Am)**
		Was a case-control design avoided?	Did the study include both positive (benign conditions) as well as healthy controls?	**(Am)**
		Did the study avoid inappropriate exclusions?	Did the study avoid inappropriate exclusions?	**(Un)**
	Index test	Were the index test results interpreted without knowledge of the results of the reference standard?	Was the index test and interpretation of data performed in standardised and reproducible fashion?	**(Am)**
		If a threshold was used, was it pre-specified?	Validation of results performed (internal or external)?	**(Am)**
	Reference standard	Is the reference standard likely to correctly classify the target condition?	Is the reference standard likely to correctly classify the target condition?	**(Un)**
		Were the reference standard results interpreted without knowledge of the results of the index test?	-	**(Om)***
	Flow and Timing	Was there an appropriate interval between index test and reference standard?	Was there an appropriate interval between index test and reference standard?	**(Un)**
		-	Were index test and reference standard performed prior to therapeutic intervention?	**(Ad)**
		Did all patients receive the same reference standard?	Did all patients receive the same reference standard?	**(Un)**
		Were all patients included in the analysis?	Were all patients included in the analysis?	**(Un)**
Applicability	Patient selection	Are there concerns that the included patients and setting do not match the review question?	Are there concerns that the included patients and setting do not match the review question?	**(Un)**

	Index test	Are there concerns that the index test, its conduct, or interpretation differs from the review question?	Are there concerns that the authors have not demonstrated suitable reproducibility and sensitivity of the chosen index test?	**(Am)**
	Reference standard	Are there concerns that the target condition as defined by the reference standard does not match the question?	Are there concerns that the target condition as defined by the reference standard does not match the question?	**(Un)**
	Am = amendment. Om = omitted. Ad = addition. Un = unchanged. *Criteria omitted as not applicable in the case of phase 1 biomarker discovery studies REVIEW QUESTION Test population: human subjects Index test(s): VOC analysis within human samples Reference standard: the accepted standard for diagnosis of cancer and/or benign disease in that field Target condition: cancer Setting: hospital, medical centre Intended use of the index test: diagnostic Patient presentation: routine investigation for symptoms of malignancy Prior testing: not applicable			

This table setting out the modified QUADAS2 scoring criteria is taken directly from a published systematic review previously done by our group at Imperial College London, relating to VOC detection. (102)

7.2 **Appendix 2: Modified QUADAS-2 assessment breakdown for each included study**

		Risk of bias				Applicability			Ref
Study	**Disease**	**Patient selection**	**Index test**	**Reference standard**	**Flow and timing**	**Patient selection**	**Index test**	**Reference standard**	
Rieder 2016	IBD	L	H	H	H	L	H	L	(118)
Hicks 2015	IBD	L	H	L	L	L	H	L	(114)
Bodelier 2015	IBD	L	L	L	?	L	L	L	(119)
Patel 2014	pIBD	?	H	L	?	L	L	L	(120)
Dryahina, 2017	IBD	H	H	H	?	?	H	H	(121)
Monastra, 2017	pIBD	L	L	L	?	L	H	L	(115)
Smolinska, 2017	pIBD	L	L	L	L	L	L	L	(122)
Walton, 2013	IBD	?	L	H	H	H	H	H	(123)
Ahmed, 2016	IBD	H	H	L	H	L	H	L	(124)
Ahmed 2013	IBD	L	L	L	?	L	L	L	(113)
Markar 2019	CRC	L	L	L	?	L	L	L	(109)
Amal, 2016	CRC	H	H	L	H	L	H	L	(117)
Altomare, 2015	CRC	H	H	L	H	H	L	L	(116)
Altomare, 2013	CRC	H	L	L	H	L	L	L	(84)
Bond, 2019	CRC	L	L	L	L	L	L	H	(125)
Kim, 2019	CRC	H	H	H	H	L	H	H	(126)
Wang, 2014	CRC	H	H	L	?	L	L	H	(127)
Wang, 2014	CRC	L	H	L	H	L	L	L	(128)
Crotti, 2016	CRC	L	L	L	?	L	L	H	(129)
Tan, 2013	CRC	H	L	?	?	L	L	L	(130)
Nishiumi, 2012	CRC	H	L	L	H	L	L	L	(131)
Cheng, 2012	CRC	H	L	H	H	L	L	H	(132)
Batty, 2015	HR polyps	H	H	L	L	H	H	H	(112)

See above, appendix 1, for detail of the components of the QUADAS-2 assessment tool

Key:
IBD = Inflammatory bowel disease
pIBD = Paediatric inflammatory bowel disease
CRC = Colorectal cancer
HR = High risk
L = low risk of bias
H = high risk of bias
? = unclear risk of bias

7.3 Appendix 3: Consent form- Non-invasive testing for gastrointestinal disease

Imperial College London

Department of Surgery & Cancer
Imperial College London
10th Floor, QEQM Building
St Mary's Hospital
Praed Street, London W2 1NY
Tel: +44(0)20 7886 2125

PARTICIPANT CONSENT FORM

Non-invasive diagnostic testing for Gastro-intestinal disease

Principal Investigator: **Professor George Hanna** PhD FRCS
Professor of Surgical Sciences / Consultant Surgeon
g.hanna@imperial.ac.uk

Please initial box

1. I confirm that I have read and understood the patient information leaflet (Version 12.0 / 02/05/2019) for the above study and have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. I understand that this part of a large study involving 3000 patients. ☐
2. I understand that my participation is voluntary and that I may withdraw consent to my samples being included in the study at any time without justifying my decision and without affecting my medical care or legal rights. ☐
3. I agree to donate my samples and allow their use in the medical research as described in the Patient Information Leaflet. ☐
4. I understand that sections of any of my medical notes may be looked at by responsible individuals from the research team or from regulatory authorities where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records that are relevant to this research. ☐
5. I understand that my samples will not be used to undertake any genetic tests. ☐
6. I understand that the samples are a gift and that I will not benefit from any intellectual property privilege that result from the use of the samples in research studies. ☐
7. I would like to be contacted if the result of any research carried out on my samples can help me during my treatment course. ☐
8. The samples which I hereby consent to donate are: a. Breath ☐
   b. Urine ☐
9. I understand I may be provided with a portable breath sensor for home use which must be returned at the end of the study period. ☐

7.4 Appendix 4: Participant Information Leaflet - Non-invasive testing for gastrointestinal disease

Imperial College London

Department of Surgery & Cancer
Imperial College London
10th Floor, QEQM Building
St Mary's Hospital
Praed Street, London W2 1NY
Tel: +44(0)20 7886 2125
Fax: +44(0)20 7413 0470

Professor George Hanna PhD FRCS
Professor of Surgical Sciences
g.hanna@imperial.ac.uk
www.imperial.ac.uk

PARTICIPANT INFORMATION LEAFLET

Non-invasive diagnostic testing for Gastro-intestinal disease

Principal Investigator:

Professor George Hanna PhD FRCS
Professor of Surgical Sciences / Consultant Surgeon
Imperial College, St Mary's Hospital, London

Co-investigators:

Mr Sheraz Markar MRCS
Clinical Research Fellow
Imperial College London

Mr Tom Wiggins MRCS
Clinical Research Fellow
Imperial College London

Mr. Sacheen Kumar MRCS
Clinical Research Fellow
Imperial College London

Dr. Tanzeela Khalid
Research Scientist
Imperial College London

Ms. Sophie Doran
Clinical Research Fellow
Imperial College London

Dr Georgia Woodfield
Clinical Research Fellow
Imperial College London

Date: 15/02/2018

Dear Sir/Madam

You are being invited to take part in the above titled research study. This information sheet explains the nature of the research being undertaken and what the process involves. Please take your time to read the following information and discuss with others if you wish. Should you require any further information or have any questions, please ask us. Take time to decide whether or not you wish to participate. Thank you for reading this information sheet.

What is the purpose of the study?
We are trying to develop a new way to diagnose cancer of the oesophagus, stomach, colon and rectum in a way that doesn't involve having an endoscopy. To do this we are collecting breath and urine samples. These samples contain metabolites, chemicals produced by cancer cells that can be analysed and may show whether cancer is present. If this analysis is successful we hope to develop a test that can detect cancer without the need for an endoscopy at an early stage when treatment is likely to be more successful. The aim is to include an overall of 3000 patients in the study to increase the accuracy of the breath test.

Why have I been chosen?
You have been chosen for this research project because you are having an endoscopy to diagnose the cause of your gastrointestinal symptoms. We want to see if new non-invasive tests on your breath and urine an also help to diagnose the cause of your symptoms and possibly replace the need for some endoscopies in the future.

Do I have to take part?
Participation in this research study is entirely voluntary. It remains your decision at all times whether or not to take part. If you choose to take part, you will be provided with a copy of this information sheet to keep and requested to sign a consent form. Should you decide at any point that you do not wish to continue with your participation in this research study, you are free to withdraw at any time. A reason for withdrawal need not be given. A decision not to take part or to withdraw at any time shall not affect the standard of care you receive throughout your treatment.

What do I have to do?
You are not required to do anything specific. In accordance with endoscopy and theatre protocol, we request for you not to eat or drink anything for at least 6 hours prior to attending the hospital. You shall undergo your investigations and treatment as planned. The collection of breath and urine will not interfere with your planned treatment nor shall it result in any change to your planned treatment. No additional invasive procedures will be undertaken for the purposes of the research study. We also request your permission to access your hospital records for the purpose of the research study only, including blood tests, radiology and pathology results. All your hospital records shall be handled with strict confidentiality in accordance with the Data Protection Act 1998.

As part of the study you will be asked to fill a Hospital Anxiety Disability Score (HADS) questionnaire. This is a validated questionnaire tool to assess anxiety levels experienced by patients undergoing endoscopic testing. The aim is to assess the psychological impact of endoscopic testing, providing the rationale of a non-invasive test like breath testing in the future. The use of the HADS score has been suggested by research and clinical psychologists.

How much of my breath and urine will be taken?
Before your procedure, you will be asked to provide two breath samples: one which stores the sample in a collection bag for analysis later and the other will require you to give multiple exhalations in a breath sampling device until a defined volume (no more than 2 Litres) has been collected. The device we will be using contains an adsorbent tube that will collect your breath samples. We expect this to take no longer than 5 minutes. You may also be asked to pass a small quantity of urine (20mL) into a container for analysis later. No extra needle pricks or vein punctures are required. The sample shall undergo analysis via our mass spectrometry instruments.

You may be provided with a hand-held breath sensor to record measurements of breath VOC before, during, or after your treatment pathway. You will be trained in how to use the device at home and given instructions on the frequency of measurements to record. The device will need to be returned at the end of the study period.

What will happen to the samples?
The machine which we will use to analyse your breath samples will destroy these once we have finished looking at them. Therefore no clinical material will remain. Any breath samples we have not looked at will be discarded. Any surplus urine samples will be stored until the end of the study should we need to re-look at these.

Will I be contacted again in the future?
You shall need to indicate on the consent form whether or not you would like to be contacted again. We shall only contact you to obtain an update on your condition. Whether or not you chose to be contacted shall not impact in any way on the standard of care that you receive; this information may be discussed with you by your Hospital Consultant.

What if something goes wrong?
We do not believe that you would be harmed by donating breath or urine samples during this study. Your treatment pathway shall remain the same irrespective of whether or not you choose to participate in this research study. If you wish to complain, or have any concerns about the way you have been approached or

treated during the course of this study, the standard National Health Service complaints procedure is available to you.

Will I receive payment for the samples that I donate to the study?
There shall be no payment or remuneration for any sample provided. We shall treat your samples as a gift to the St. Mary's Hospital Upper GI Research Group, London and you would therefore relinquish any interest in the samples provided.

Who is the organising and funding body of the tissue bank and study?
Division of Surgery, Department of Surgery & Cancer, Imperial College London, St. Mary's Hospital, Praed Street, London, W2 1NY.
.
Who has reviewed the ethical considerations of the study?
This study has been reviewed and given favourable ethical opinion for conduct in the NHS by Camden & Islington REC.

Thank you again for taking the time to read this information leaflet. Your participation in this research is most appreciated.

I have some more questions, with whom may I get in contact?

Professor George B Hanna PhD FRCS
Professor of Surgical Sciences / Consultant Upper GI Surgeon
Department of Surgery & Cancer
10th Floor, QEQM Building
St Mary's Hospital, London
Praed street,
London W2 1NY
Tel: 0207 886 2124
Fax: 0207 886 1810
E-mail: g.hanna@imperial.ac.uk

Mr Sheraz Markar
Clinical Research Fellow
Department of Surgery & Cancer
Room 1089
10th Floor, QEQM Building
St Mary's Hospital, London
Praed street,
London W2 1NY

7.5 Appendix 5: Participant Information Leaflet- Non-invasive testing for colorectal disease

Imperial College London

Department of Surgery & Cancer
Imperial College London
10th Floor, QEQM Building
St Mary's Hospital
Praed Street, London W2 1NY
Tel: +44(0)20 7886 2125
Fax: +44(0)20 7413 0470

Professor George Hanna PhD FRCS
Professor of Surgical Sciences
g.hanna@imperial.ac.uk
www.imperial.ac.uk

PARTICIPANT INFORMATION LEAFLET

Non-invasive diagnostic testing for colorectal disease

IRAS ID 223074

Principal Investigator:

Professor George Hanna PhD FRCS
Professor of Surgical Sciences / Consultant Surgeon
Imperial College, St Mary's Hospital, London

Co-investigators:

Dr Georgia Woodfield
Clinical Research Fellow
Imperial College London

Mr Geng-Ping Lin
Clinical Research Fellow
Imperial College London

Ms. Sophie Doran
Clinical Research Fellow
Imperial College London

Mr Piers Boshier
Surgical registrar
Imperial College London

Dear Sir/Madam

You are being invited to take part in the above titled research study. This information sheet explains the nature of the research being undertaken and what the process involves. Please take your time to read the following information and discuss with others if you wish. Should you require any further information or have any questions, please ask us. Take time to decide whether or not you wish to participate. Thank you for reading this information sheet.

What is the purpose of the study?

We are trying to develop a new way to detect diseases of the colon and rectum, in a way that doesn't involve having an endoscopy. We are predominantly focusing on the detection of colorectal cancer and polyps, but are also interested in the presence of other colonic diseases. To do this we are collecting breath, and in some cases urine and colonic air samples. These samples contain metabolites, which are chemicals produced by cancer cells and polyps, that can be analysed and may show whether these conditions are present. If this analysis is successful we hope to develop a test that can detect polyps and detect cancer without the need for an endoscopy at an early stage when treatment is likely to be more successful.

Why have I been chosen?

You have been chosen for this research project because you are having an endoscopy, having an operation, or having a consultation with a doctor. We want to see if new non-invasive tests on your breath and/or urine can help diagnose the cause of symptoms or predict the outcomes of your investigations. This could possibly replace the need for some endoscopies in the future.

Do I have to take part?

Participation in this research study is entirely voluntary. It remains your decision at all times whether or not to take part. If you choose to take part, you will be provided with a copy of this information sheet to keep and requested to sign a consent form. Should you decide at any point that you do not wish to continue with your participation in this research study, you are free to withdraw at any time. A reason for withdrawal need not be given. A decision not to take part or to withdraw at any time shall not affect the standard of care you receive throughout your treatment.

What do I have to do?

As part of the study you will asked to give a sample of your breath, and possibly a sample of your urine and sample of colonic air (taken during the endoscopy procedure). These collections will occur whilst you are attending your planned appointment. If you are attending a consultation, you attend as usual. If you are attending endoscopy or for an operation, in accordance with endoscopy and theatre protocol, we request for you not to eat or drink anything for at least 6 hours prior to attending the hospital. You shall undergo your investigations and treatment as planned. The collection of breath and/or urine and/or colonic air for the study will not interfere with your planned treatment nor shall it result in any change to your planned treatment. No additional invasive procedures will be undertaken for this research.

We also request your permission to access your hospital records for the purpose of the research study only, including blood tests, radiology and pathology results. All data will be anonymised and all hospital records shall be handled with strict confidentiality in accordance with the Data Protection Act 1998. Data will be kept for 10 years after the end of the study, unless a participant withdraws from the study, in which case study data will be destroyed.

How much of my breath and urine/colonic air will be taken?

Before your procedure or consultation, you will be asked to provide a breath sample. This involves breathing normally into a breath sampling device until a defined volume (no more than 2 Litres) has been collected. The device we will be using contains adsorbent tubes that collect your breath samples. We may also ask you to breathe into a breath-collection bag for analysis later. This involves one exhalation into the bag. In some cases we may also ask you to breathe directly into a larger breath sampling device machine, which will involve one breath only per measurement. We expect this to take no longer than 5 minutes. You may also be asked to pass a small quantity of urine (20mL) into a container for analysis later. Patients who are undergoing a colonoscopy may also be asked to consent to have up to 500ml of air from their colon collected. The air is suctioned up via the colonoscope during the procedure. Suctioning of air is a routine part of the examination, but for this study we would be collecting it for analysis rather than letting it escape. The sample(s) will all undergo analysis via our mass spectrometry instruments.

What will happen to the samples?

The machine which we will use to analyse your breath samples will destroy these once we have finished looking at them. Therefore no clinical material will remain. Any breath or air samples we have

not looked at will be discarded. Any surplus urine samples will be stored until the end of the study should we need to re-look at these.

Will I be contacted again in the future?

You should indicate on the consent form whether or not you would like to be contacted in the event that results could help you during you treatment course. Whether or not you chose to be contacted shall not impact in any way on the standard of care that you receive; this information may be discussed with you by your Hospital Consultant.

What if something goes wrong?

We do not believe that you will be harmed by donating breath or urine or colonic air samples during this study. Your treatment pathway shall remain the same irrespective of whether or not you choose to participate in this research study. If you are harmed by taking part in this research project, there are no special compensation arrangements. If you are harmed due to someone's negligence, then you may have grounds for a legal action. Regardless of this, if you wish to complain, or have any concerns about any aspect of the way you have been treated during the course of this study then you should immediately inform the Investigator (G Hanna, see below). The normal National Health Service complaints mechanisms are also available to you. If you are still not satisfied with the response, you may contact the Imperial AHSC Joint Research Compliance Office.

Will I receive payment for the samples that I donate to the study?

There shall be no payment or remuneration for any sample provided. We shall treat your samples as a gift to the St. Mary's Hospital Upper GI Research Group, London and you would therefore relinquish any interest in the samples provided.

Who is the organising and funding body of the tissue bank and study?

Division of Surgery, Department of Surgery & Cancer, Imperial College London, St. Mary's Hospital, Praed Street, London, W2 1NY. The study is also receiving funding from the Rosetrees Trust (charity number 1119976).

Who has reviewed the ethical considerations of the study?

This study has been reviewed and given favourable ethical opinion for conduct in the NHS by Essex REC.

Thank you again for taking the time to read this information leaflet. Your participation in this research is most appreciated.

I have some more questions, with whom may I get in contact?

Professor George B Hanna PhD FRCS
Professor of Surgical Sciences / Consultant Upper GI Surgeon
Department of Surgery & Cancer
10th Floor, QEQM Building
St Mary's Hospital, London, Praed street, London W2 1NY
Tel: 0207 886 2124
E-mail: g.hanna@imperial.ac.uk

Dr Georgia Woodfield
Clinical Research Fellow
Department of Surgery & Cancer
Room 1089, 10th Floor, QEQM Building
St Mary's Hospital, London, Praed street, London W2 1NY
E-mail: g.woodfield@imperial.ac.uk

Mr Geng-Ping Lin
Clinical Research Fellow
Department of Surgery & Cancer
Room 1089, 10th Floor, QEQM Building
St Mary's Hospital, London, Praed street, London W2 1NY
E-mail: g.lin15@imperial.ac.uk

Alternatively, you can seek impartial advice from the Patient Advice and Liaison Service (PALS), at PALS, Ground Floor QEQM, St Mary's Hospital, 41 Praed Street, London W2 1NY, Tel: +44(0)2078867777, Fax: +44(0)2078861753

Lastly, the trust R&D provide a third point of contact: Ms Ruth Nicholson AHSC Joint Research Compliance Office, 510B, 5th Floor Lab Block, Charing Cross Hospital, Fulham Palace Road, London, W6 8RF Tel: +44 (0) 203 311 0212 Fax: +44 (0) 203 311 0212 r.nicholson@imperial.ac.uk

7.6 Appendix 6: Consent form- Non-invasive testing for colorectal disease

Imperial College London

Department of Surgery & Cancer
Imperial College London
10th Floor, QEQM Building
St Mary's Hospital
Praed Street, London W2 1NY
Tel: +44(0)20 7886 2125

PARTICIPANT CONSENT FORM

Non-invasive diagnostic testing for colorectal disease
IRAS ID 223074

Principal Investigator: **Professor George Hanna** PhD FRCS
Professor of Surgical Sciences / Consultant Surgeon
g.hanna@imperial.ac.uk

Please initial box

1. I confirm that I have read and understood the information leaflet for the above study and have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐
2. I understand that my participation is voluntary and that I may withdraw consent to my samples being included in the study at any time without justifying my decision and without affecting my medical care or legal rights. ☐
3. I agree to donate my samples and allow their use in the medical research as described in the Patient Information Leaflet (version 3.0, 24/4/2017). ☐
4. I understand that sections of any of my medical notes may be looked at by responsible individuals from the research team or from regulatory authorities where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records that are relevant to this research. ☐
5. I understand that my samples will not be used to undertake any genetic tests. ☐
6. I understand that the samples are a gift and that I will not benefit from any intellectual property privilege that result from the use of the samples in research studies. ☐
7. I would like to be contacted if the result of any research carried out on my samples can help me during my treatment course. ☐
8. The samples which I hereby consent to donate are:
   a. Breath ☐
   b. Urine ☐
   c. Colonic air ☐

_____	_____	_____
Name of Patient	Signature	Date
_____	_____	_____
Researcher	Signature	Date

1 copy to be given to the participant, 1 copy in medical notes and original to be filed in trial site file.

Patient consent form COBRA Version 2.0 15/2/17

7.7 Appendix 7: Baseline clinical data form

COBRA Study- endosocopy patients only **Patient study ID:**.....
To be completed after completion of consent form **Age:**..... **BMI**.....

Date:..... **Time:**..... **How many hours ago did you last eat?** .....

Bowel prep taken (tick multiple) : Senna □ Citramag □
Moviprep □ Klean prep □ Phosphate enema □ Other □ None □

Gender: Female □ Male □ Other □

Reason for colonoscopy? BCSP (screening) □ Symptoms- 2WW □ Symptoms- urgent □
Symptoms- routine □ Surveillance (eg. prev polyps/IBD/FH of cancer) □ Re-scope for polyp removal □

Ethnic Origin (tick more than one for mixed)
White □ Black / African / Caribbean / Black British □ Asian / Asian British □
Arab □ Other ethnicity □

Symptoms (if relevant): can tick multiple

Change in bowel habit □	Diarrhoea □	Constipation □
Black stool □	Rectal bleeding □	Weight loss □
Abdominal pain □	Reduced appetite □	Passing mucus □
Pain on defecation □	Incomplete emptying □	Mass in abdomen □

Other:

Past Medical history

Previous colorectal cancer □	Previous other cancer □	Previous known polyps □
Ulcerative colitis □	Crohn's disease □	Other IBD □
Coeliac disease □	Past bowel resection □	Barrett's oesophagus □
Recent antibiotics □	Diabetes □	High blood pressure □
Known heart disease □	Kidney disease □	Liver disease □
Asthma □	COPD/emphysema □	Other lung disease □

Type of past cancer or past bowel surgery (if applicable):

Medications:

Omeprazole/lansoprazole (Losec, Prilosec, Prevacid) □		Ranitidine (zantac) □
Aspirin/Ibuprofen/Nurofen □	Clopidogrel (Plavix) □	Blood thinner e.g warfarin □
Laxative □	Antibiotics □	Immunosuppressants □

Other medications (if applicable):

Smoking:
Never □ Ex □ Current □ Number of cigarettes/day.....
Alcohol:
Never □ Ex □ Current □ Number of units/week.....

Family history of colorectal cancer in First Degree Relative (youngest if multiple):
No □ Yes FDR <50 □ Yes FDR>50 □

Colonoscopy report (If there are complicated/lengthy findings, you can staple report here and no need to fill out):

Normal □

Polyp(s) □

Cancer(s) □

Diverticular disease □

Inflammatory bowel disease (UC/Crohn's/ indeterminate) □

Infectious colitis including C.difficile □

Colitis of uncertain type □

Haemorrhoid □

Stricture □

Fistula □

Other abnormality (describe below)

If unsure if the lesion is a polyp or a cancerous polyp- fill out as per polyp

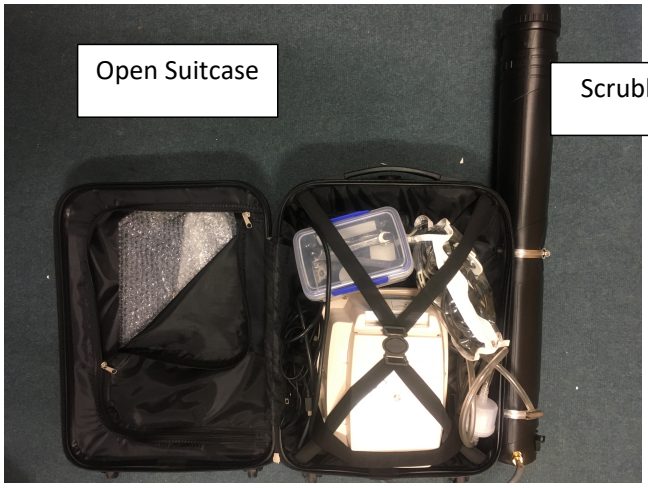
Polyps/cancers : Circle the answer for each polyp/cancer, or for multiple polyps/cancers copy correct answer into box from selection :

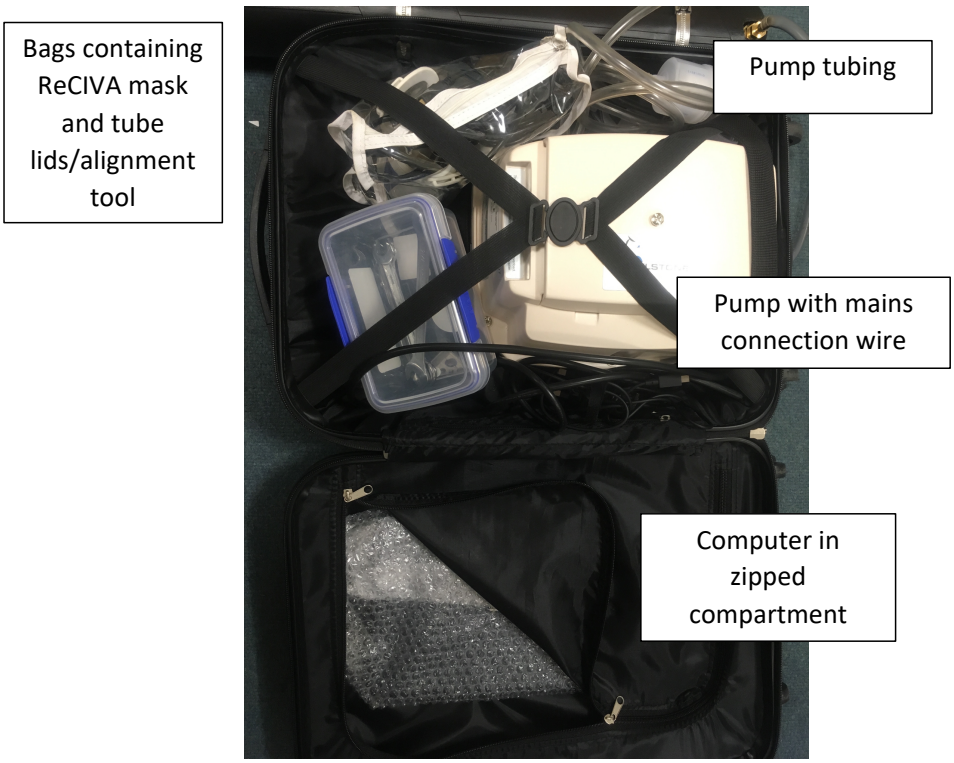
	Site	Size in mm	Type of polyp	Likely cancerous? Y/N/?	Histology (to be filled out later)
Lesion 1	-Anus -Rectum -Sigmoid -Desc. Colon -Splenic flexure -Transverse colon -Hepatic flexure -Asc. colon -Caecum -Ileocaecal Valve -Terminal ileum		Stalked Raised (sessile) Flat Depressed		
Lesion 2					
Lesion 3					
Lesion 4					
Lesion 5					
Lesion 6					
Lesion 7					
Lesion 8					
Lesion 9					
Lesion 10					

7.8 Appendix 8: ReCIVA® List of tasks

ReCIVA Sequential Instructions, for COBRA study- May 2017

Contents of the ReCIVA suitcase





Phase 1- Preparation before the patient arrives

Assemble

1. Ensure that the Study ID (written on the sample bag containing the tubes) is copied across to the top of the consent form and patient details forms.

   The study ID is written like this on each sample bag: CO_ _0001.

   Please fill in the second letter of the first name and second letter of the surname to make the full anonymous study ID.

---

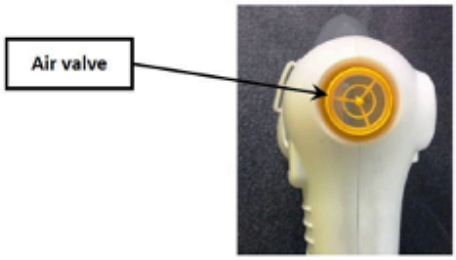
2. Wear gloves



---

3. Inspect the air valve (yellow part in picture below) to check the membrane in the valve isn't damaged and is sitting correctly.





---

4. Collect the unopened Mask and 4 Sorbent tubes

---

5. **Prepare the tubes**

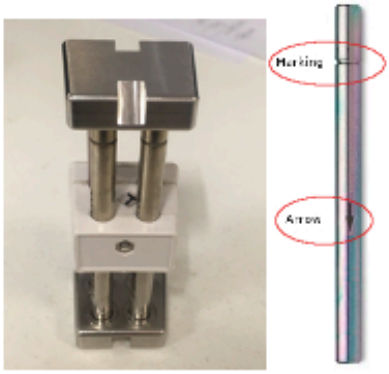
   a. Using the spanner tool, remove the brass caps from the ends of the sorbent tubes. Insert the tubes with the grooves facing upwards into the alignment tool.
   NB. T=Top





b. Seal the 4 tubes together with the metal lids. Tubes are safe to be kept like this until a patient comes to be sampled.

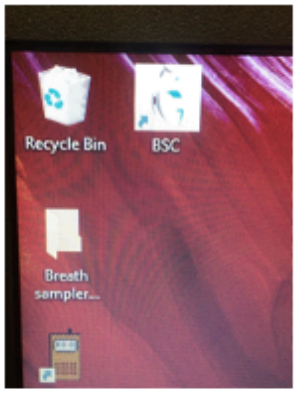




6. Prepare the computer

a) Plug the computer into the mains. The password is written on the keyboard. Click on "BSC" on the screen and the ReCIVA programme will open. It will say "please connect mask" on screen.





b) Connect the handheld ReCIVA to the computer using the black connector cord via the USB port.



The computer screen will show that it has detected the device and a moving display will appear. You will see the option to "start breath collection" at the bottom.

7. Prepare the pump.

a) The pump is strapped into the suitcase and does not need to be removed. The scrubber is the long cylindrical device encased in a plastic tube connected to the suitcase.
Connect the pump tubing to the scrubber using the pump connector. You will need to unscrew the lid of the scrubber to expose the connector. It simply pushes in to connect.



b) Connect the pump lead to the mains. You will hear it start. Turn it off at the plug until you need to use it.
Once the rest of the equipment is removed from the suitcase, you can zip the suitcase back up and keep the pump hidden, as you will not need to adjust this again.

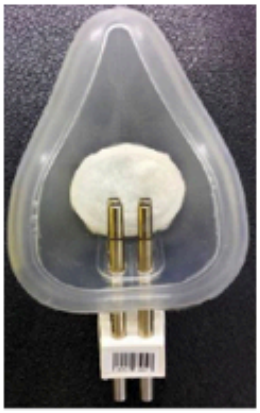
c) Connect the end of the pump tubing to the hand held ReCIVA.



Phase 2: You are now ready to sample a patient

8. Firstly, explain the study and the test, and offer a patient information leaflet. Ask the patient to sign a consent form and give them the opportunity to ask questions.
9. If they agree to participate, open a new rubber mask, so that the patient can see that it is clean.
    a. Remove the top lid from the 4 tubes. Insert the tubes into the 4 holes in the bottom of the mask from below. The grooves in the tubes should be facing upwards so that they are inside the mask.

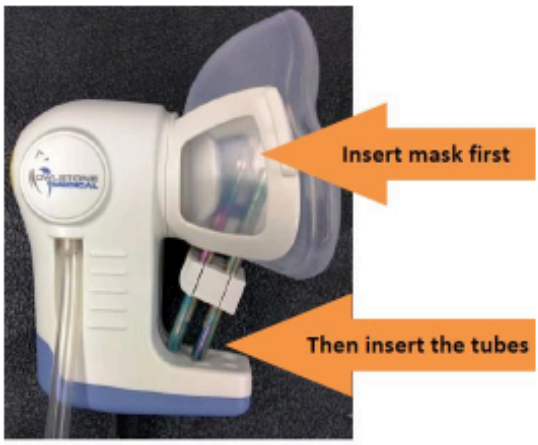
    b. You then need to remove the bottom lid, so that the tubes are dangling freely.





Sorbent tubes in mask

10. Insert the mask first into the ReCIVA sampler, then push the tubes into the holes

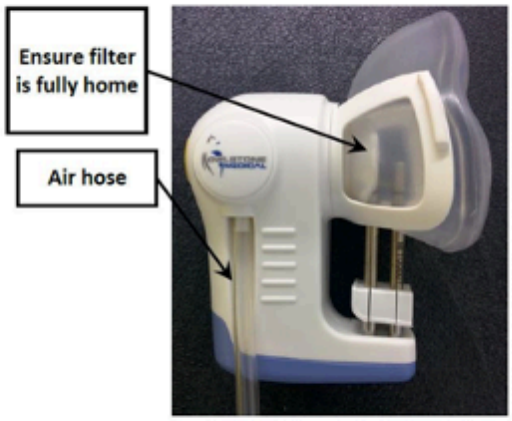


Mask inserted –
NOTE - tubes not inserted yet.

11. If you are finding it difficult to push the tubes down into the holes by hand, you can use the square metal tube lid to push the tubes down into position.



12. Connect the air hose from the pump to the attachment on the ReCIVA device, if it not already connected (see step 7c). Flow is around 30 Litres/min. Remember to turn the pump on at the plug.



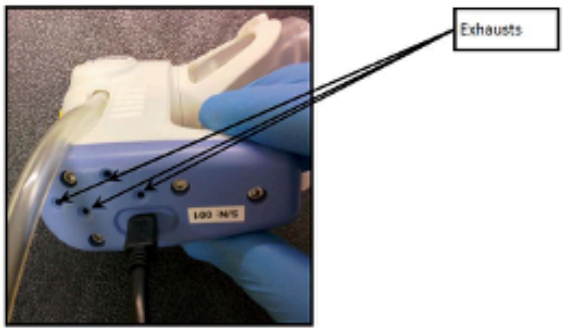
13. If not already done, connect ReCIVA sampler to the computer through the USB port place at the bottom part of ReCIVA sampler (see step 6b).
If the ReCIVA is already connected, I suggest closing down the programme and reopening it (BSC) if the programme has been running idle for a long time. This will ensure that it is ready to go.

14. Final check-
Ensure pump is on at the plug
Ensure BCS Programme is loaded on the computer
Ensure mask is set up

15. Press " start breath collection " on the computer screen. An error message will appear. This always happens. Click "continue". The computer will then wait until the patient starts breathing, and you will see a moving trace appear.

16. Ask patient to put the mask in place, forming a seal with the skin, and start to breathe normally. Either nasal or mouth breathing is acceptable. It will say "learning" for a few seconds in the bottom of the screen, and then "collecting".

NOTE: make sure the patient does not block Exhausts.



17. Check and correct (if needed) the mask position if there are any leaks.

18. Explain to patients to breathe normally:

- <<Please breathe normally until the end of the process, I will let you know when the sample collection is over, it will take approximately 2 or 3 minutes, >>
- Encourage the patient to keep a good seal between their face and the mask.

IMPORTANT NOTE: Patients can talk during the collection, but this will slow the collection.
If the patient removes the mask, this is not a problem; the sampler will stop sampling automatically and will restart when breathing commences again. However, encourage the patient not to do this, as the collection will take longer.
If the patient is able, this can be a good time to fill in the patient details (CRF) form.
If the patient is distressed – stop the test

USE

End

19. When software indicates that the sampling is completed, the test is finished

20. Thank patient/discharge/debrief

Disassemble

Phase 3: Labelling samples and Disassembling the equipment

21. Pull out the mask and the Sorbent tubes from the ReCIVA. Remove and dispose of the mask.

22. Pull all the tubes out and seal them by screwing the brass nuts onto the ends of the tubes, using the spanner device if needed. Ensure these are tight and cannot be pulled off. The bottom of the brass cap should reach the groove if it is on correctly.



23. Insert the 4 sealed TD tubes into the sample packs that they came in. You can ensure that it is the correct bag by checking the barcode numbers that are written on the label of the bag.

24. Make sure the unique patient ID is complete by adding the second letter of the first name then the second letter of the second name.

25. Write in the practice name, date and time of sampling onto the label that is affixed onto the plastic sample bag.

26. Put the whole bag into the Tupperware box, and seal it so that it is airtight. All sample bags can go in the same Tupperware box back to the lab for processing.

27. Ensure that the patient information form and consent form are placed in the opaque folder. This folder should be transported with the samples, as it will need to be stored in a locked room at St Mary's.

28. To pack away the equipment, it is relatively straight-forward to unplug everything and put it all back in the suitcase.
One tip: TO DISCONNECT PUMP TUBING FROM SCRUBBER- this involves pushing down on the cylindrical connector so that the fitting is released. Do not pull it out! It will not release until you push the fitting down on the scrubber.

7.9 Appendix 9: COBRA protocol

Imperial College London

Imperial College Healthcare NHS Trust

CLINICAL PROTOCOL

Full Title	Non-invasive testing for the diagnosis and assessment of colorectal disease
Short Title	Colorectal breath analysis (COBRA)
Sponsor	Imperial College London Ruth Nicholson Room 501B, 5th Floor Lab Block, Charing Cross Hospital Fulham Palace Road London W6 8RF
Chief Investigator	Professor George Hanna Head of Division of Surgery Imperial College Healthcare NHS Trust
Principal Investigator	Professor George Hanna
Management Group	Dr Georgia Woodfield, Mr Geng-Ping Lin, Mr Piers Boshier, Mr Andrea Romano

TABLE OF CONTENTS

1. Signature Page 3
2. Summary 4
3. Introduction 5
3.1 Background 5
3.2 Rationale for study 7
4. Study Aims 8
4.1 Aims 8
5. Methodology 8
5.1 Inclusion Criteria 8
5.2 Exclusion Criteria 8
5.3 Study Design and Plan 8
6. Study Procedures 10
6.1 Study Timetable 10
6.2 Recruitment and Informed Consent Procedures 10
7. Statistical Considerations 10
8. Ethical Considerations 11
8.1 Declaration of Helsinki and Good Clinical Practice 11
8.2 Ethical Committee Review 11
8.3 Consent 11
9. Safety Considerations and Reporting 12
9.1 Adverse Event 12
9.2 Serious Adverse Event (SAE) 12
9.3 Unexpected Adverse Event 12
9.4 Expected Adverse Events 12
9.5 Reporting Unexpected Adverse Events 12
9.5.1 Assessment of Intensity 12
9.6 Urgent Safety Measures 13
10. Data Handling and Record Keeping 13
11. Specimen anonomisation and storage 13
12. Specimen analysis 14
13. Devices and Techniques 15
13.1 Devices 15
14. Monitoring and Auditing 15
15. Study Committees 15
15.1 Study Steering Committee 15
16. Finance and Funding 15
17. Indemnity 15
18. Dissemination of Research Findings 15
19. References 16

1. SIGNATURE PAGE

Chief Investigator Agreement

The clinical study as detailed within this research protocol (), or any subsequent amendments will be conducted in accordance with the Research Governance Framework for Health & Social Care (2005), the World Medical Association Declaration of Helsinki (1996) and the current applicable regulatory requirements and any subsequent amendments of the appropriate regulations.

Chief Investigator Name:

Chief Investigator Site:

Signature and Date:

Principal Investigator Agreement

The clinical study as detailed within this research protocol (), or any subsequent amendments will be conducted in accordance with the Research Governance Framework for Health & Social Care (2005), the World Medical Association Declaration of Helsinki (1996) and the current applicable regulatory requirements and any subsequent amendments of the appropriate regulations.

Principal Investigator Name:

Principal Investigator Site:

Signature and Date:

2. SUMMARY

Short Title	Colorectal cancer breath analysis (COBRA)
Methodology	Cohort Observational & Laboratory study
Research Site	St Mary's Hospital Charing Cross Hospital St Mark's Hospital St George's Hospital Homerton University Hospital Chelsea and Westminster Hospital West Middlesex Hospital Royal Marsden Hospital
Aims	Determine the diagnostic accuracy of the exhaled breath and urine test for the prediction of colorectal disease. To assess the acceptability of using breath and urine testing
Main Inclusion Criteria	Any patient who: 1. Patients aged 18 - 90 years 2. Patients seen in secondary care with symptoms of colorectal disease or referred for a lower gastrointestinal endoscopy as part of their routine clinical investigation programme 3. Patients able to understand and retain the information provided, thereby being able to give informed consent for inclusion in this study
Number of Participants	Target 2000
Endpoints	*Primary endpoint* • Once samples have been collected from 2000 patients.
Study Duration	April 2017 to April 2022

3. INTRODUCTION

3.1 Background

Colorectal cancer (CRC) is the 2nd commonest cause of cancer death in the UK. Survival rates are among the lowest in Europe (1) and 26% of patients present as an emergency, frequently having visited their GP before diagnosis (2-4). If diagnosed early, CRC survival rates exceed 90% (5). Whilst, reducing delay in diagnosis is therefore considered key to improving outcomes [6], initiatives to reduce delay have thus far been largely unsuccessful (6, 7). In order to improve the diagnosis of CRC, efforts should be made to determine how the risk of CRC can be assessed using sensitive primary care tests that are: valid; easy to perform, and; well tolerated by patients.

CRC screening programmes rely on detection of early-stage cancer using the faecal occult blood test (FOBT) with colonoscopy in those patients with a positive test. Trials have shown that this screening method reduced CRC mortality but not incidence rates. Even if UK CRC incidence rates remain constant, as they have for the past 30 years, with an aging population the number of cases and treatment cost are set to double in the next 50 years. The Bowel Scope programme is also being rolled out across the UK currently, aiming to perform flexible sigmoidoscopy on every person at the age of 55. Invasive screening reduces mortality but is not without risk to patients and pressurises an already over burdened healthcare service.

Risk of CRC in symptomatic patients can be assessed by different methods:

(I) *Flexible sigmoidoscopy and colonoscopy*: The NHS currently has insufficient trained endoscopists to perform the number of procedures that would be required if flexible sigmoidoscopy were used as a first-line test for symptoms with predictive values below 3%.

(II) *Guaiac FOBT* has good sensitivity (87-98% in CRC detection), but highly variable and often unsatisfactory specificity (13-79%), requiring the repetition of the test on multiple stool samples. To date, the FOBT is neither recommended nor available for use as an intermediate test (8-11).

(III) *Faecal immunochemical testing (FIT)* requires sampling only a single stool. It has been extensively tested; and is now used by many CRC screening programmes with very low prevalence rates of CRC (12). Four FIT systems are fully automated, and provide a quantitative measure of haemoglobin, allowing selection of a threshold of positivity to fit specific circumstances. As a result the research data available on sensitivity and specificity for CRC is based on small numbers of cancers. The few studies that have included adequate numbers of cancers have used older qualitative tests with a fixed threshold for positivity and were not automated. Nevertheless, the data suggest that, depending on the selected threshold for positivity, sensitivity for CRC varies between 35% and 86% with specificity between 85% and 95% (8, 9). However, there are no data on the sensitivity of the newer quantitative test for early-stage cancers.

(IV) Multi-target stool DNA testing: When compared with the FIT, in a large multicentre study, the test showed better specificity (92 *vs.* 73%), but lower sensitivity (90 *vs.* 96%) (13).

At Imperial College London, we have developed a breath test for the detection of oesophagogastric adenocarcinoma from biomarker discovery to clinical trials. The assessment of volatile organic compound (VOCs) within exhaled by selected ion flow tube mass spectrometry (SIFT-MS) (14) was followed by the construction and validation of a diagnostic model using a panel of 9 VOCs in 225 patients with an area under the ROC curve (±SE) of 0.92±0.01 and 0.87±0.03 for the model and validation subsets respectively (15). We also performed cross-platform chemical validation using GC-MS. A NIHR-funded multicentre blind validation study on 396 patients (16) showed a sensitivity of 86% and specificity of

80%. In parallel with this work we have also investigated the exhalation kinetics and molecular drivers of VOCs in oesophagogastric adenocarcinoma (17-19).

In colorectal cancer we performed a prospective pilot study to identify discriminative VOCs (n=150); a diagnostic validation study (n=110); a clinical prognostic study following surgery and after tumour recurrence (n=60); cross platform chemical validation using GC-MS (n=30). The diagnostic model had a sensitivity of 85% and specificity of 84% for cancer diagnosis. Following surgery the biomarker normalised to control level, but levels significantly increased with recurrence (20). The volatile biomarkers for oesophago-gastric and colorectal cancers are different with no overlap.

Volatile Organic Compounds (VOCs)

VOCs emitted from the human body have been of interest to researchers for several decades. In 1971, Pauling et al (21), reported that breath and urine contained approximately 250 and 280 VOCs respectively in normal human subjects. VOCs are traditionally measured in food industry and counter-terrorism (22-24).

The analysis of exhaled breath for the non-invasive detection and monitoring of disease is an attractive and evolving field of clinical research. Routine clinical applications of breath testing include the alcohol breathalyser, ^{13}C urea breath test for H. pylori, nitric oxide in asthma, and hydrogen/methane test for bacteria overgrowth (25-29).

Alteration in VOC production in patients with cancer has been postulated to relate to oxygenation of cell membrane-based polyunsaturated fatty acids resulting from genetic and/or protein mutations within tumour cells and the increased relative prevalence of reactive oxygen species (30, 31). These VOCs produced from the cancer site within the body, travel within blood to reach pulmonary alveoli where they are exhaled, permitting an objective quantitative measurement (32). There have been early reports using VOCs to detect lung, breast, bladder, prostate and gastric cancers (33, 34). Reports on colorectal cancer had small patient cohorts and lacked validation (32, 35, 36).

Due to the complex composition of human breath a number of sophisticated mass spectrometer based technologies have been employed for the detection of exhaled VOCs including Gas Chromatography coupled with Mass Spectrometry (GC/MS), Proton Transfer Reaction-Mass Spectrometry (PTR-MS), Selected Ion Flow Tube-Mass Spectrometry (SIFT-MS) and Ion Mobility Spectrometry (IMS). These instrumental methods all have their strengths and limitations for applications in breath analysis. We intend to use multiple analytical platforms in our proposed work as these techniques are complementary and will allow us to acquire more robust findings.

Thermal-desorption (TD) coupled to Gas Chromatography Mass Spectrometry (GC/MS) is a widely adopted method for measuring breath VOCs. Although this method requires the pre-concentration of breath volatiles prior to analysis, it allows for the comprehensive identification and semi-quantification of VOCs in unknown complex mixtures such as breath. It can separate and detect hundreds of compounds from a breath sample, allowing for VOC profiling and has therefore been used in the study of disease biomarkers for lung cancer (34, 37), breast cancer (33), colorectal cancer (32) and liver diseases (38, 39).

Selective Ion Flow Tube Mass Spectrometry (SIFT-MS) permits online, real-time VOC

quantification (40). SIFT-MS has been utilised in the study of VOCs in breath and urine from patients with conditions including cystic fibrosis and bladder cancer (41, 42). The SIFT-MS technique allows real-time detection and quantification of VOCs within biological samples such as exhaled breath without any sample preparation minimising diagnostic delay.

Proton-Transfer-Reaction Time-of-Flight Mass Spectrometry (PTR-ToF-MS) is a relatively new instrumental tool that is gaining interest in breath research applications. Similarly to SIFT-MS, Proton-Transfer-Reaction Mass Spectrometry (PTR-MS) is a direct injection-based technology, obviating problems related to pre-concentration and separation steps typical of GC-MS. The most important advantage guaranteed by the ToF mass analyzer over quadrupole-based instruments (GC-MS and SIFT-MS) is represented by the high time resolution. In a ToF analyser a complete mass spectrum is generated at each ionisation event and spectra can be recorded at a frequency of 1 Hz or lower. This improves the accuracy of measurement during *in vivo* studies; moreover, full spectral scans can be performed without compromising sensitivity, thus enabling simultaneous targeted and untargeted analysis.

The use of multi-platform analytical methods will enable both compound identification and quantification which are very important in the search and validation of biomarkers.

3.2 Rationale

Colonoscopy remains the gold-standard investigation for the assessment of patients with lower gastrointestinal symptoms and considered at risk of colorectal cancer. Colonoscopy is an expensive investigation that is uncomfortable for the patient and not without important risks including visceral perforation and bleeding.

A breath test is a non-invasive investigation. The use of a breath test as a triage investigation in symptomatic patients could identify high-risk patients who should be referred for colonoscopy at an earlier stage, increase the proportion of appropriate referrals from primary care, and improve NICE guidelines which may currently miss particularly younger patients (43). If a GP is presented with a patient with gastrointestinal symptoms that do not prompt referral under NICE guidelines, he/she would not need to watch-and-wait to see if symptoms worsen but could offer the test immediately. The GP would order a breath test in much the same way as routine blood tests, with a single breath collection assessing for individual or multiple cancers. A nurse can perform the test and send breath samples to a regional laboratory for analysis. A positive result would warrant immediate referral. A negative test would permit the GP to reassure the patient and offer retesting if symptoms persist. Furthermore if we could develop a breath test that can discriminate colonic adenomas, the test would have further a potential application within the screening programme.

A breath test could also be used to detect recurrence in patients post-resection for their colorectal cancer, allowing earlier detection and limiting invasive testing in those who are free of recurrence.

Previous pilot study data for colorectal cancer detection by our group produced a diagnostic model with a sensitivity of 85% and specificity of 84% for colorectal cancer. The next stage of this research and the subject of this current proposal is to assess the influence of lower gastrointestinal diseases on VOC exhaled breath and urine. The external validation of this VOC breath model in a large multi-centre study will provide this risk-stratification tool for future clinical practice.

4. STUDY AIMS

4.1 Aims

The primary aim of this study is to determine the diagnostic accuracy of the exhaled breath and urine test for the prediction of colorectal disease.

Secondary aims are:
(i) Refine the current diagnostic model using technologies that allow identification of additional biomarkers at a lower detection level for CRC.
(ii) Confirm the diagnostic accuracy of the test in a new and independent group of patients who have CRC compared to subjects with benign diseases or normal colon.
(iii) Confirm the prognostic accuracy of the test in detecting tumour recurrence after surgical removal of CRC.
(iv) Profile VOCs in patients with colonic adenomas and construct a diagnostic model.
(v) Profile VOCs from colonic air in a subset of patients, for use as a comparison against breath testing in order to improve understanding of underlying pathological mechanisms.
(vi) To assess the acceptability of using breath and urine testing

5. METHODOLOGY

5.1 Inclusion Criteria

Any patient who is:

- ≥ 18 years old and below 90 years of age, AND:
- undergoing a lower gastrointestinal endoscopy as part of their routine clinical care OR:
- seen in another hospital clinical setting with symptoms of lower gastrointestinal disease
- able to provide informed written consent

5.2 Exclusion Criteria

Any patient who:

- Lacks capacity or is unable to provide informed consent.
- Any patient below 18 years of age or over 90 years of age.

5.3 Study Design and Plan

Patients recruited from the endoscopy unit will have been referred by their responsible clinician to undergo a lower gastrointestinal endoscopy (which involves a telescope test to visualise the colon). Others may be attending for colonoscopy as part of the bowel cancer screening programme or as part of routine surveillance. When patients attend the endoscopy unit, a member of the research team will discuss the research project with the patient, explain that the study involves sampling of their breath and urine while they are waiting for their endoscopy. Patients may also be asked to consent to the use of the colonic air collected during colonoscopy for analysis. Patients will then be provided with a patient

information leaflet, (see Participant Information Leaflet version 3.0 24/04/2017). Should the patient agree to participate they will be asked to sign a written consent form (see Consent Form Version 2.0 15/02/2017).

Patients recruited from secondary care will be approached either before or after a routine appointment that they will be attending as part of their clinical care. This could include general colorectal clinics, pre-operative colorectal clinics, oncology clinics, theatres or on hospital wards.

The project will be explained to them and they will be provided with an information leaflet. If they agree to participate they will be asked to sign a written consent form in order to provide breath and urine samples for the research project. Consent for colonic air sampling is not applicable to the majority of secondary care patients, only to those having a colonoscopy. No additional invasive procedures will be undertaken for the purposes of this study. A subset of patients with known cancer who are due to have a cancer resection may be asked to provide additional breath and urine samples 6–8 weeks following surgical resection of their tumour and/or if their disease should recur.

Patients attending for colonoscopy will be fasted for a minimum of six hours prior to their breath sample collection. This is normal routine practice for patients undergoing endoscopy and therefore will not alter their treatment pathway. All breath samples will be retrieved prior to endoscopy. In other secondary care settings patients will not be specifically asked to fast prior to giving breath or urine samples but note will be made of how long they have fasted for. Patients will be asked to give a number of exhalations into a breath sampling device until a defined volume of breath has been sampled. This involves performing normal tidal breathing whilst wearing a face mask, and takes around 2 minutes on average. Patients will also be asked to perform a single deep nasal inhalation followed by complete exhalation via their mouth into a secure GastroCHECK steel breath bag (500 mL) via a 1 mL Luer-Lok syringe (Terumo Europe, Leuven, Belgium). Patients at St Mary's Hospital will also be asked to breathe directly into a mass spectrometer machine. The mass spectrometer is a self-contained mobile device which can be easily moved between clinical areas within St Mary's Hospital. Sampling by this method will involve a single deep nasal inhalation followed by complete exhalation via their mouth into a plastic cylinder connected to a tube leading to the mass spectrometer itself. We expect the breath sampling process to take no longer than 5-10 minutes in total. Patients will be requested to pass urine into a standardised 60ml urine specimen vial, which will be immediately sealed. 20ml of urine will be aliquoted into a standard 60ml specimen vial.

Of patients attending for colonoscopy, a subset will be asked to consent to have air from their colon collected. This air comes up from the colon through the colonoscope suction channels. Suctioning of air is a routine part of the examination, but for this study we will be collecting it for analysis. Collection is done via a small suction pump connected to a fine bore tube that is passed through the suction channel of the colonoscope. Sampling in this way will add no more than a few minutes (<4 minutes) on to the procedure time.

The VOCs from the breath/colonic air will be trapped onto adsorbent tubes inserted into the breath taking device. In the case of colonic air sampling, an adsorbent tube is attached to the pump end of the fine bore disposable suction tube. After sampling the adsorbent tube is removed from the device, capped and transported back to the laboratory at St Mary's Hospital London for analysis. Samples will be transported by courier or by a member of the research team. The sample(s) shall undergo analysis via our mass spectrometry instruments, at St Mary's Hospital London.
Breath samples collected in steel bags will be immediately disposed of following analysis within 6 hours of the sample being taken. Breath VOCs sampled on adsorbent tubes will be

analysed within 48 hours of collection. The sample constituents are destroyed within the mass spectrometer for detection meaning that no clinical material will remain after analysis by GC-MS. Urine samples will be stored within a registered tissue bank (REC Ref no. 04/Q0403/119, custodian Prof G Hanna) in accordance with the Human Tissue Act and disposed of following analysis.

This study has been specifically designed to ensure that any disruption to patient's routine care and management is kept to an absolute minimum.

6. PROCEDURES

6.1 Study Timetable

January 2017 – March 2017
Research & Development and ethical approval.

April 2017 – March 2022
Participant recruitment and sample collection. Laboratory analysis of samples, write up and dissemination of results.

6.2 Recruitment and Informed Consent Procedures

A member of the research team will identify and approach patients when they attend for a lower gastrointestinal endoscopy. All patients attending for colonoscopy within the specified age range are potentially eligible for this study. If the patient is in agreement, a member of the research team will explain the research project and provide a written information leaflet (see Participant Information Leaflet version 3.0 24/04/2017). Patients will be given adequate time to decide if they wish to participate and should they agree they will complete a written consent form (see Consent Form Version 2.0 15/02/2017).

A member of the research team will identify patients when they attend for a clinic appointment by discussion with the Consultant clinician responsible for the clinic. A member of the research team will approach patients when they attend for their clinic or pre-assessment appointment prior to planned surgery. Patients will have the research project explained to them and will be provided with a written information leaflet (see Participant Information Leaflet version 3.0 24/04/2017). Patients will be provided with adequate time to decide if they wish to participate and should they agree they would complete a written consent form (see Consent Form Version 2.0 15/02/2017).

All of the above would occur at one of the patient's routine clinical attendances and would not require any additional patient attendances.

7. STATISTICAL CONSIDERATIONS

Both univariate and multivariate data analysis techniques will be applied to the results to 1) identify VOC components with the best discriminating ability between the groups and 2) to develop a multivariate discriminant analysis model.

Mann-Whitney U test will be used to compare the measured metabolite concentrations between the study groups. A p value <0.05 will be taken as the level to indicate statistical significance.

Non-parametric (Kruskal-Wallis) ANOVA test will be used to compare the measured metabolite levels between the study groups. Statistically significant metabolites will be selected using False Discovery Rate (FDR) adjusted p value threshold such that less than 5% of discovered candidate markers are false positives (i.e. q-value <0.05).
The FDR strategy will be based on Benjamini–Hochberg procedure.

Receiver operating characteristic (ROC) curves will be used to determine the accuracy of a diagnostic test in classifying those with and without colorectal disease. The metabolite concentrations that exhibit significant differences (q<0.05) between the cancer and positive controls (including adenomatous polyps) will be included as variables for the proposed diagnostic test. The linear discriminant analysis based on multivariate logistic regression will be subsequently used to derive a combination of the metabolite concentrations for sample classification. The optimum threshold for group discrimination will be calculated to achieve the highest possible classification accuracy. The predictive capacity of the model will be assessed on independent sample set, as described below.

Validation
The Breath and Urine metabolite model for the prediction of colorectal disease developed in the previous studies will be used to assign risk of disease in a prospective cohort. Investigators undertaking metabolite sampling and analysis will be blinded to the results of the endoscopy. Comparison of predicted disease risk and actual endoscopy findings will then be made, and the overall diagnostic accuracy (sensitivity, specificity, positive and negative predictive value, ROC curve analysis) for this non-invasive diagnostic investigation will be determined. Colonic air VOC profiles will be used for validation of paired breath results, and to gain further understanding of the mechanism of VOC production in humans.

8. ETHICAL CONSIDERATIONS

8.1 Declaration of Helsinki and Good Clinical Practice

The study will conform to the spirit and letter of the declaration of Helsinki and in accordance with the Good Clinical Practice Guidelines

8.2 Ethical Committee Review

The study protocol will be used when approval has been obtained from a UK Research Ethics Committee. The study must be submitted for Site Specific Assessment (SSA) at each participating NHS Trust. The Chief Investigator will require a copy of the Trust R&D approval letter before accepting participants into the study. The study will be conducted in accordance with the recommendations for physicians involved in research on human subjects adopted by the 18th World Medical Assembly, Helsinki 1964 and later revisions. Additionally, in order to sample breath from any patient who is having a colonoscopy via the UK Bowel Cancer Screening Programme (BCSP) pathway, there also has to be specific approval from the UK BCSP Research Advisory committee. Formal approval has been gained for an initial 100 BCSP patients to be approached for this study, as a pilot, with a view to further approval after this period (BCSP ID189, approval given on 18/1/17). Further approval will therefore be sought for the recruitment of additional BCSP patients, after 100 patients have been approached, as per BCSP panel request.

8.3 Consent

Consent to enter the study must be sought from each participant only after a full explanation has been given, an information leaflet offered and time allowed for consideration. Signed

participant consent should be obtained. The right of the participant to refuse to participate without giving reasons must be respected. All participants are free to withdraw at any time from the protocol without giving reasons and without prejudicing further treatment.

9. SAFETY CONSIDERATIONS AND REPORTING

This is not a study of an investigational medicinal product and therefore all untoward occurrences will be adverse events rather than adverse reactions.

9.1 *Adverse Event*

An adverse event is defined as any untoward medical occurrence affecting a participant. The terms '*mild, moderate or severe*' are used to describe the intensity of a specific event or reaction. An event would be considered '*serious*' based upon participant/event outcome or action criteria as defined below.

9.2 *Serious Adverse Event (SAE)*

A serious adverse event (SAE) will be considered any untoward medical occurrence/effect that:

- Results in death
- Is life-threatening
- Requires hospitalisation or prolongation of existing inpatient's hospitalisation.
- Results in persistent or significant disability or incapacity.
- Is otherwise considered medically significant by the investigator

'Life threatening' refers to an event in which the subject was at risk of death at the time of the event; it does not refer to an event which hypothetically may have caused death if it had been more severe.

9.3 *Unexpected Adverse Event*

An unexpected adverse event is considered an adverse event of which the nature and severity is not consistent with the expected consequences of sample collection.

9.4 *Expected Adverse Events*

We do not believe any risks or adverse events are associated with patients providing samples of urine and breath or in the consent process for these procedures. All samples will be disposed of following analysis. All hospital records will be handled with strict confidentiality in accordance with the Data Protection Act 1998. All other risks will be in accordance with usual clinical practice. These include risk of endoscopic perforation, patient discomfort during endoscopy, and local bruising following intravenous cannulation. There are no other serious adverse events caused by sample collection and tissue biopsy.

9.5 *Reporting Unexpected Adverse Events*

Given the nature of this investigation, it is unlikely that any will occur. Nonetheless, should an unexpected adverse events occur, the investigators will make an assessment of severity and report the event to the Research Ethics Committee and the Sponsor as appropriate. Usually, this will mean within 24 hours in the case of severe or moderate events, and within 14 days for mild events.

9.5.1 *Assessment of Intensity*

Mild: The subject is aware of the event or symptom, but the event or symptom is easily tolerated.
Moderate: The subject experiences sufficient discomfort to interfere with or reduce their usual level of activity.
Severe: Significant impairment of functioning; the subject is unable to carry out usual activities and/or the subject's life is at risk from the event.

9.6 *Urgent Safety Measures*

Should it be necessary to undertake urgent safety measures to ensure the safety and protection of the participants the CI will take these measures immediately. The CI would inform the sponsor and Main research Ethics Committee of this event immediately via telephone and in writing within 3 days in the form of a substantial amendment.

10. DATA HANDLING AND RECORD KEEPING

Information related to study participants will remain confidential and be managed in accordance with the Data Protection Act, NHS Caldecott Principles, The Research Governance Framework for Health and Social Care, and the conditions of Research Ethics Committee Approval.

All case record forms will be held in the site file, which will be kept in a locked drawer in a locked office at the study site. They will be destroyed after 10 years. All digital data will be stored on a password-protected NHS computer on site at St Mary's Hospital, under the control of the Chief Investigator. Data will be erased securely after 10 years.

11. SPECIMEN ANONOMISATION AND STORAGE

Samples will be anonymised with all patients being assigned a study number that will be added to the specimen during storage. Patients would not be identifiable from this study number. These study numbers will be linked to the digital database stored on a password-protected NHS computer on site at St Mary's Hospital, under the control of the Chief Investigator.

Breath samples collected in steel bags will be disposed of immediately following analysis. Breath/colonic air VOCs sampled on to adsorbent tubes will be released by thermal desorption into the analytical instrument where they will be destroyed during mass spectrometric detection. Urine samples will be stored within -80C freezers used solely for the storage of human samples within a secure laboratory within the research department at St Mary's hospital. The only people who will have access to this tissue bank are key members of the research team. These samples will be added to Professor Hanna's current tissue bank for which local ethical approval is already in place for sample collection at St Mary's Hospital (REC Ref no. 04/Q0403/119). Urine samples will be stored until study completion. Analysed samples will be destroyed by incineration following analysis and surplus samples will be stored in the current tissue bank until study completion in case a repeat analysis is required.

12. SPECIMEN ANALYSIS

Selective Ion Flow Tube Mass Spectrometry (SIFT–MS)
SIFT-MS permits online, real-time Volatile Organic Compound (VOC) quantification. SIFT-MS has been utilised in the study of VOCs in breath and urine from patients with conditions including cystic fibrosis and bladder cancer. The principle of SIFT-MS is selected precursor ions are formed in a microwave discharge source and are selected according to their mass-to-charge ratio, m/z, by a mass filter and injected into a helium carrier gas where they are convected as a thermalised swarm along a flow tube. H_3O^+, NO^+, O_2^+ precursor ions are used to ionize the trace gases in an air sample that is introduced into the helium at a known flow rate, these ions selectively ionise VOCs present within the sample resulting in characteristic product ions. By measuring the count rate of both precursor ions and the characteristic product ions at the downstream detection system, a real-time quantification is achieved, realising the absolute concentration of trace and volatile compounds at the parts-per-billion by volume or parts-per-million by volume. The SIFT-MS technique allows real-time detection and quantification of Volatile Organic Compounds within biological samples such as exhaled breath without any sample preparation minimising diagnostic delay. This is particularly advantageous within the clinical environment where samples can be retrieved and real-time VOC measurements made with negligible concern for sample degradation.

Thermal Desorption (TD) and Gas Chromatography Mass Spectrometry (GC/MS)
Thermal desorption coupled to GC/MS is one of the most widely used techniques for breath biomarker discovery work for a number of reasons. Firstly, VOCs in exhaled breath are present at very low concentrations of parts-per-million (ppm) to parts-per-trillion (ppt) and pre-concentration is often necessary to allow for detection. This can be achieved by adsorption onto sorbent beds (tubes packed with tenax, carbon molecular sieve, or graphitized carbon) followed by release through thermal desorption. The coupling of a thermal desorber to a GC/MS system allows for the trapped VOCs to be released and rapidly injected on to the GC column for analysis. GC/MS is a very well established analytical technique that allows for the successful separation, identification, and semi-quantification of trace VOCs in unknown complex mixtures such as breath. This technique has allowed for the identification of over 800 compounds in breath (44). Although this technique does not allow for real-time analysis it provides a much more comprehensive analysis of the VOC profile. It will complement results obtained with SIFT-MS and can allow for cross-platform validation.

Proton-Transfer-Reaction Time-of-Flight Mass Spectrometry (PTR-ToF-MS)
Similarly to SIFT-MS, Proton-Transfer-Reaction Mass Spectrometry (PTR-MS) is a direct injection-based technology, obviating problems related to pre-concentration and separation steps typical of GC-MS. In PTR-MS the ion beam generated at the source does not undergo selection by means of a mass filter as in SIFT-MS so the detection limits are on average two orders of magnitude lower than those obtained with SIFT-MS. However, equipped with a Time-of-Flight (ToF) mass analyzer, a higher mass resolution is achieved, signals represented by a single mass peak in a quadrupole-generated spectrum, can be resolved in two or more distinct mass peaks, generally resulting in an improved discrimination performance. In addition to this, accurate masses can be determined with at least two decimal digits; this means that the data generated by PTR-ToF-MS, especially when integrated with GC-MS, can be used for compound identification and thus permit the discovery of new biomarkers. The most important advantage guaranteed by the ToF mass analyzer over quadrupole-based instruments (as those in GC-MS and SIFT-MS) is represented by the high time resolution. In a ToF analyzer a complete mass spectrum is generated at each ionisation event and spectra can be recorded at a frequency of 1 Hz or lower. This improves the accuracy of measurement during *in vivo* studies; moreover, full

spectral scans can be performed without compromising sensitivity, thus enabling simultaneous targeted and untargeted analysis.

13. DEVICES AND TECHNIQUES

13.1 *Devices*

A device that collects and concentrates breath VOCs on adsorbent tubes in one easy step will be used to collect breath (Owlstone's CE-marked hand-held breath sampler, ReCIVA, Owlstone Ltd, Cambridge, UK). This device has been designed for easy patient use, it contains a pump of filtered air to ensure clean air sampling, it allows for the concentration of a defined volume, and is more economical than other sampling methodologies available. A steel bag will also be used for the collection of breath. For direct sampling into the mass spectrometer, the patient blows into a disposable cardboard tube attached to the inlet to the machine, discarded between each patient. All other devices are those used within the patient's routine clinical practice including endoscopic equipment and universal sampling containers. These will all be used in accordance with routine clinical practice.

14. MONITORING AND AUDITING

The study will be monitored and audited by the Sponsor, in accordance with usual protocols.

15. STUDY COMMITTEES

a. *Study Steering Committee*

This will comprise of Professor George Hanna, Professor Wendy Atkins, Professor Brian Saunders, Dr Jonathan Hoare

16. FINANCE AND FUNDING

Equipment utilised for sample collection will be funded as part of Professor Hanna's research funding. Laboratory analysis of the biological specimens (breath, colonic air and urine samples) obtained will also be funded through Professor Hanna's research group with external funding from HCA Hospitals. A grant from the Rosetrees Trust (ref: A1361) has already been approved to fund an analytical chemist to work on this study.

17. INDEMNITY

The nominated sponsor of our research study is Imperial College London, Research and Development Department, Room 501B, Laboratory Block, Charing Cross Hospital, Fulham Palace Road, London, W6 8RF. Delegated responsibilities will be assigned to the NHS trusts taking part in this study.
Imperial College London holds negligent harm and non-negligent harm insurance policies which apply to this study. Medical co-investigators will also be covered by their own medical defence insurance for non-negligent harm.

18. DISSEMINATION OF RESEARCH FINDINGS

Results from this study will be presented in a peer-reviewed journal. Authorship will follow international guidelines.

19. References

1. De Angelis R, Sant M, Coleman MP, Francisci S, Baili P, Pierannunzio D, et al. Cancer survival in Europe 1999-2007 by country and age: results of EUROCARE-5 -- a population-based study. The Lancet Oncology.15(1):23-34.
2. Elliss-Brookes L, McPhail S, Ives A, Greenslade M, Shelton J, Hiom S, et al. Routes to diagnosis for cancer - determining the patient journey using multiple routine data sets. Br J Cancer. 2012;107(8):1220-6.
3. Public Health England. National Cancer Intelligence Network: Routes to diagnosis 2015 [Available from: http://www.ncin.org.uk/publications/routes_to_diagnosis.
4. Sheringham JR, Georghiou T, Chitnis XA, Bardsley M. Comparing primary and secondary health-care use between diagnostic routes before a colorectal cancer diagnosis: Cohort study using linked data. Br J Cancer. 2014;111(8):1490-9.
5. UK Cancer Research. Bowel cancer survival statistics - by stage. 2014 [Available from: http://www.cancerresearchuk.org/cancer-info/cancerstats/types/bowel/survival/bowel-cancersurvival-.
6. Peacock O, Clayton S, Atkinson F, Tierney GM, Lund JN. 'Be Clear on Cancer': the impact of the UK National Bowel Cancer Awareness Campaign. Colorectal Disease. 2013;15(8):963-7.
7. Thorne K, Hutchings HA, Elwyn G. The effects of the Two-Week Rule on NHS colorectal cancer diagnostic services: A systematic literature review. BMC Health Services Research. 2006;6(1):43.
8. Allison JE, Sakoda LC, Levin TR, Tucker JP, Tekawa IS, Cuff T, et al. Screening for Colorectal Neoplasms With New Fecal Occult Blood Tests: Update on Performance Characteristics. Journal of the National Cancer Institute. 2007;99(19):1462-70.
9. Allison JE, Tekawa IS, Ransom LJ, Adrain AL. A Comparison of Fecal Occult-Blood Tests for Colorectal-Cancer Screening. New England Journal of Medicine. 1996;334(3):155-60.
10. Imperiale TF, Ransohoff DF, Itzkowitz SH, Turnbull BA, Ross ME. Fecal DNA versus Fecal Occult Blood for Colorectal-Cancer Screening in an Average-Risk Population. New England Journal of Medicine. 2004;351(26):2704-14.
11. Lieberman DA, Harford WV, Ahnen DJ, Provenzale D, Sontag SJ, Schnell TG, et al. One-Time Screening for Colorectal Cancer with Combined Fecal Occult-Blood Testing and Examination of the Distal Colon. New England Journal of Medicine. 2001;345(8):555-60.
12. Benson VS, Atkin WS, Green J, Nadel MR, Patnick J, Smith RA, et al. Toward standardizing and reporting colorectal cancer screening indicators on an international level: The international colorectal cancer screening network. International Journal of Cancer. 2012;130(12):2961-73.
13. Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, et al. Multitarget Stool DNA Testing for Colorectal-Cancer Screening. New England Journal of Medicine. 2014;370(14):1287-97.
14. Kumar S, Huang J, Abbassi-Ghadi N, Spanel P, Smith D, Hanna GB. Selected ion flow tube mass spectrometry analysis of exhaled breath for volatile organic compound profiling of esophago-gastric cancer. Anal Chem. 2013;85(12):6121-8.
15. Kumar S, Huang J, Abbassi-Ghadi N, Mackenzie HA, Veselkov KA, Hoare JM, et al. Mass Spectrometric Analysis of Exhaled Breath for the Identification of Volatile Organic Compound Biomarkers in Esophageal and Gastric Adenocarcinoma. Annals of surgery. 2015;262(6):981-90.
16. Markar SR, Lagergren J, Hanna GB. Research protocol for a diagnostic study of non-invasive exhaled breath analysis for the prediction of oesophago-gastric cancer. BMJ Open. 2016;6(1).
17. Adam M. Role of microbiome in volatile production in oesophageal adenocarcinoma, PhD: Imperial College London; 2015-2018.

18. Antonowicz S. Aldehyde dysfunction in oesophageal adenocarcinoma-PhD: Imperial College London; 2013-2016.
19. Wiggins T. Phenol and aromatic acid dysfunction in oesophageal adenocarcinoma, PhD. Imperial College London; 2013-2016.
20. Markar S. Use of volatile organic compounds in the diagnosis of gastrointestinal cancers, PhD : Imperial College London; 2013-2016.
21. Pauling L, Robinson AB, Teranishi R, Cary P. Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. Proceedings of the National Academy of Sciences of the United States of America. 1971;68(10):2374-6.
22. Molgaard B, Viitanen AK, Kangas A, Huhtiniemi M, Larsen ST, Vanhala E, et al. Exposure to airborne particles and volatile organic compounds from polyurethane molding, spray painting, lacquering, and gluing in a workshop. International journal of environmental research and public health. 2015;12(4):3756-73.
23. Smith PA, Sheely MV, Hakspiel SJ, Miller S. Volatile organic compounds produced during irradiation of mail. AIHA journal : a journal for the science of occupational and environmental health and safety. 2003;64(2):189-95.
24. Xiao Z, Liu S, Gu Y, Xu N, Shang Y, Zhu J. Discrimination of cherry wines based on their sensory properties and aromatic fingerprinting using HS-SPME-GC-MS and multivariate analysis. Journal of food science. 2014;79(3):C284-94.
25. Atreja A, Fu AZ, Sanaka MR, Vargo JJ. Non-invasive testing for Helicobacter pylori in patients hospitalized with peptic ulcer hemorrhage: a cost-effectiveness analysis. Digestive diseases and sciences. 2010;55(5):1356-63.
26. Gahleitner F, Guallar-Hoyas C, Beardsmore CS, Pandya HC, Thomas CP. Metabolomics pilot study to identify volatile organic compound markers of childhood asthma in exhaled breath. Bioanalysis. 2013;5(18):2239-47.
27. Harding P, Field PH. Breathalyzer accuracy in actual law enforcement practice: a comparison of blood- and breath-alcohol results in Wisconsin drivers. Journal of forensic sciences. 1987;32(5):1235-40.
28. Petrone P, Sarkisyan G, Fernandez M, Coloma E, Akopian G, Ortega A, et al. Small intestinal bacterial overgrowth in patients with lower gastrointestinal symptoms and a history of previous abdominal surgery. Archives of surgery (Chicago, Ill : 1960). 2011;146(4):444-7.
29. Robroeks CM, van Berkel JJ, Jobsis Q, van Schooten FJ, Dallinga JW, Wouters EF, et al. Exhaled volatile organic compounds predict exacerbations of childhood asthma in a 1-year prospective study. The European respiratory journal. 2013;42(1):98-106.
30. Kneepkens CM, Lepage G, Roy CC. The potential of the hydrocarbon breath test as a measure of lipid peroxidation. Free radical biology & medicine. 1994;17(2):127-60.
31. Toyokuni S. Molecular mechanisms of oxidative stress-induced carcinogenesis: From epidemiology to oxygenomics. IUBMB Life. 2008;60(7):441-7.
32. Altomare DF, Di Lena M, Porcelli F, Trizio L, Travaglio E, Tutino M, et al. Exhaled volatile organic compounds identify patients with colorectal cancer. The British journal of surgery. 2013;100(1):144-50.
33. Phillips M, Cataneo RN, Ditkoff BA, Fisher P, Greenberg J, Gunawardena R, et al. Prediction of breast cancer using volatile biomarkers in the breath. Breast cancer research and treatment. 2006;99(1):19-21.
34. Phillips M, Gleeson K, Hughes JM, Greenberg J, Cataneo RN, Baker L, et al. Volatile organic compounds in breath as markers of lung cancer: a cross-sectional study. Lancet. 1999;353(9168):1930-3.
35. Altomare DF, Di Lena M, Porcelli F, Travaglio E, Longobardi F, Tutino M, et al. Effects of Curative Colorectal Cancer Surgery on Exhaled Volatile Organic Compounds and Potential Implications in Clinical Follow-up. Annals of surgery. 2015;262(5):862-6; discussion 6-7.
36. Amal H, Leja M, Funka K, Lasina I, Skapars R, Sivins A, et al. Breath testing as potential colorectal cancer screening tool. Int J Cancer. 2016;138(1):229-36.

37. Filipiak W, Sponring A, Filipiak A, Ager C, Schubert J, Miekisch W, et al. TD-GC-MS analysis of volatile metabolites of human lung cancer and normal cells in vitro. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2010;19(1):182-95.
38. Dadamio J, Van den Velde S, Laleman W, Van Hee P, Coucke W, Nevens F, et al. Breath biomarkers of liver cirrhosis. Journal of chromatography B, Analytical technologies in the biomedical and life sciences. 2012;905:17-22.
39. Khalid TY, Costello BDL, Ewen R, White P, Stevens S, Gordon F, et al. Breath volatile analysis from patients diagnosed with harmful drinking, cirrhosis and hepatic encephalopathy: a pilot study. Metabolomics. 2013;9(5):938-48.
40. Spanel P, Smith D. Selected ion flow tube mass spectrometry for on-line trace gas analysis in biology and medicine. European journal of mass spectrometry (Chichester, England). 2007;13(1):77-82.
41. Spanel P, Smith D. Progress in SIFT-MS: breath analysis and other applications. Mass spectrometry reviews. 2011;30(2):236-67.
42. Spanel P, Smith D, Holland TA, Al Singary W, Elder JB. Analysis of formaldehyde in the headspace of urine from bladder and prostate cancer patients using selected ion flow tube mass spectrometry. Rapid communications in mass spectrometry : RCM. 1999;13(14):1354-9.
43. The National Institute for Health and Care Excellence. Suspected cancer: recognition and referral 2015 [Available from: https://www.nice.org.uk/guidance/NG12.
44. de Lacy Costello B, Amann A, Al-Kateb H, Flynn C, Filipiak W, Khalid T, et al. A review of the volatiles from the healthy human body. Journal of breath research. 2014;8(1):014001.

7.10 Appendix 10: Mass spectra and NIST matches for the esters

VOCs 1, 8, 9 and 12 were tentatively identified as propyl propionate, allyl acetate, a similar overlapping ester to allyl acetate, and methyl 2-butynoate, based on the following NIST library matches, see Figure 7-1 to Figure 7-3.

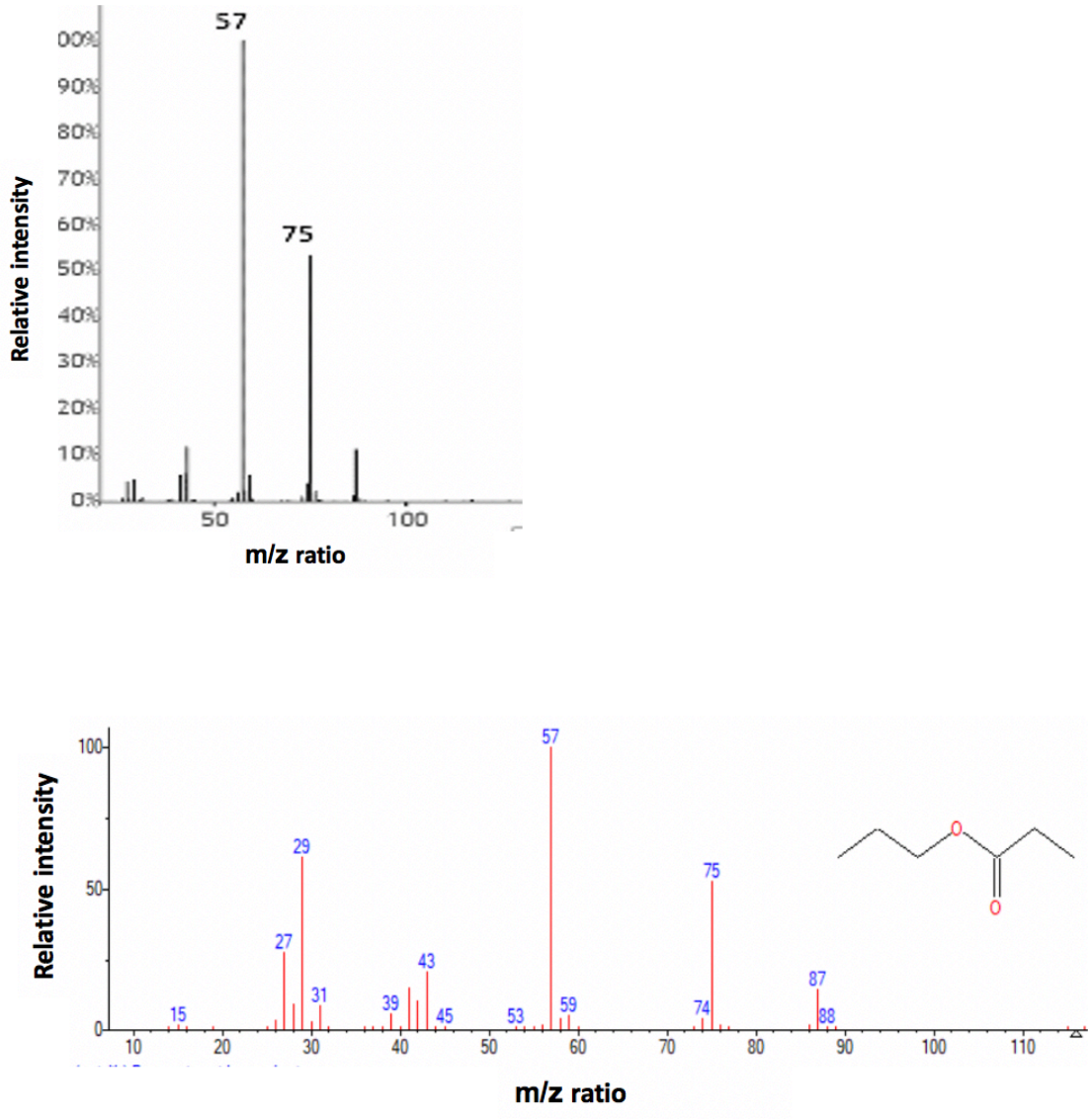


Figure 7-1 Obtained mass spectrum for VOC 1 compared to the mass spectrum of propyl propionate from the NIST database.

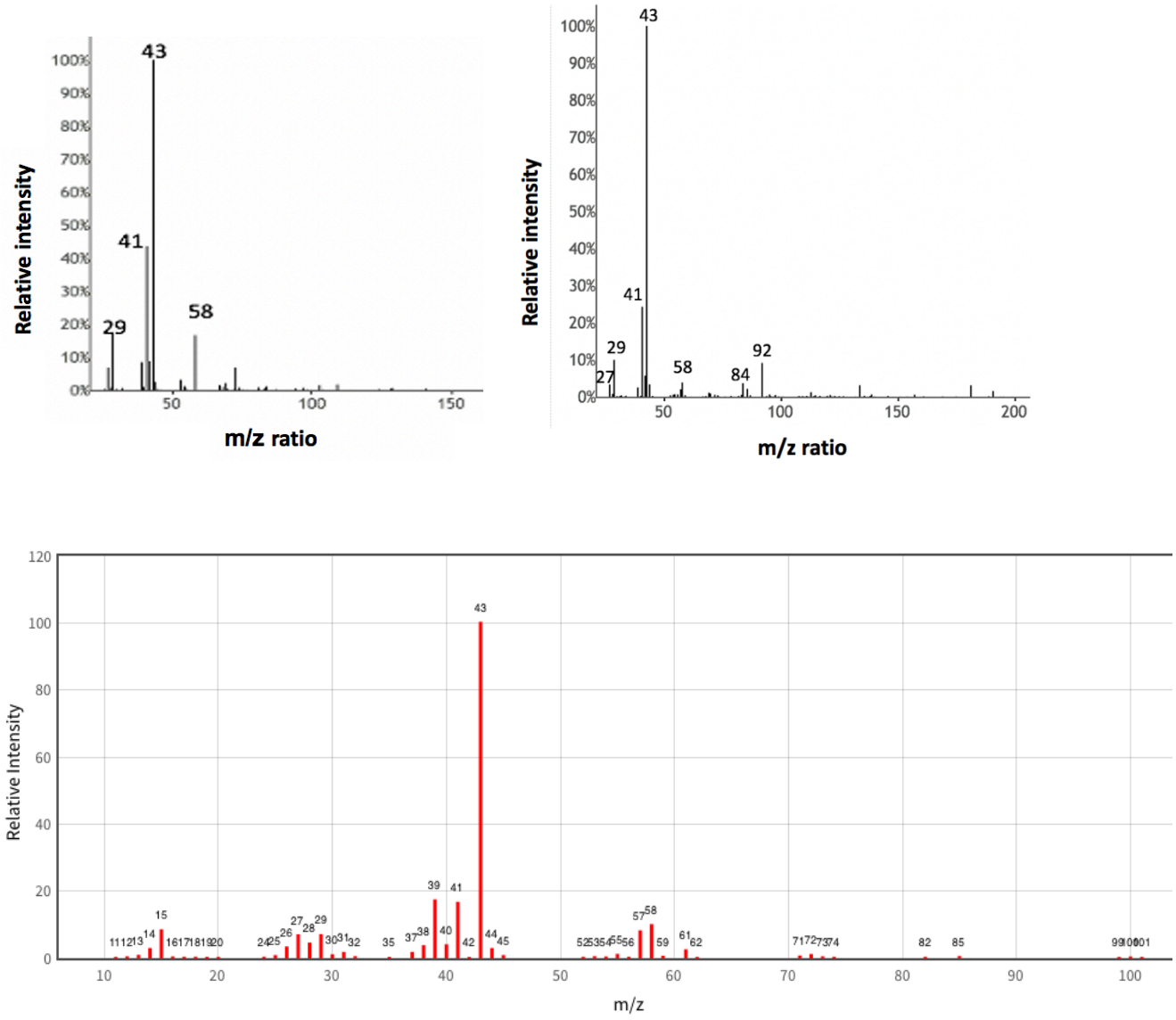


Figure 7-2 Obtained mass spectrum for VOC 8 (left) and VOC 9 (right) compared to the mass spectrum of allyl acetate from the NIST database.

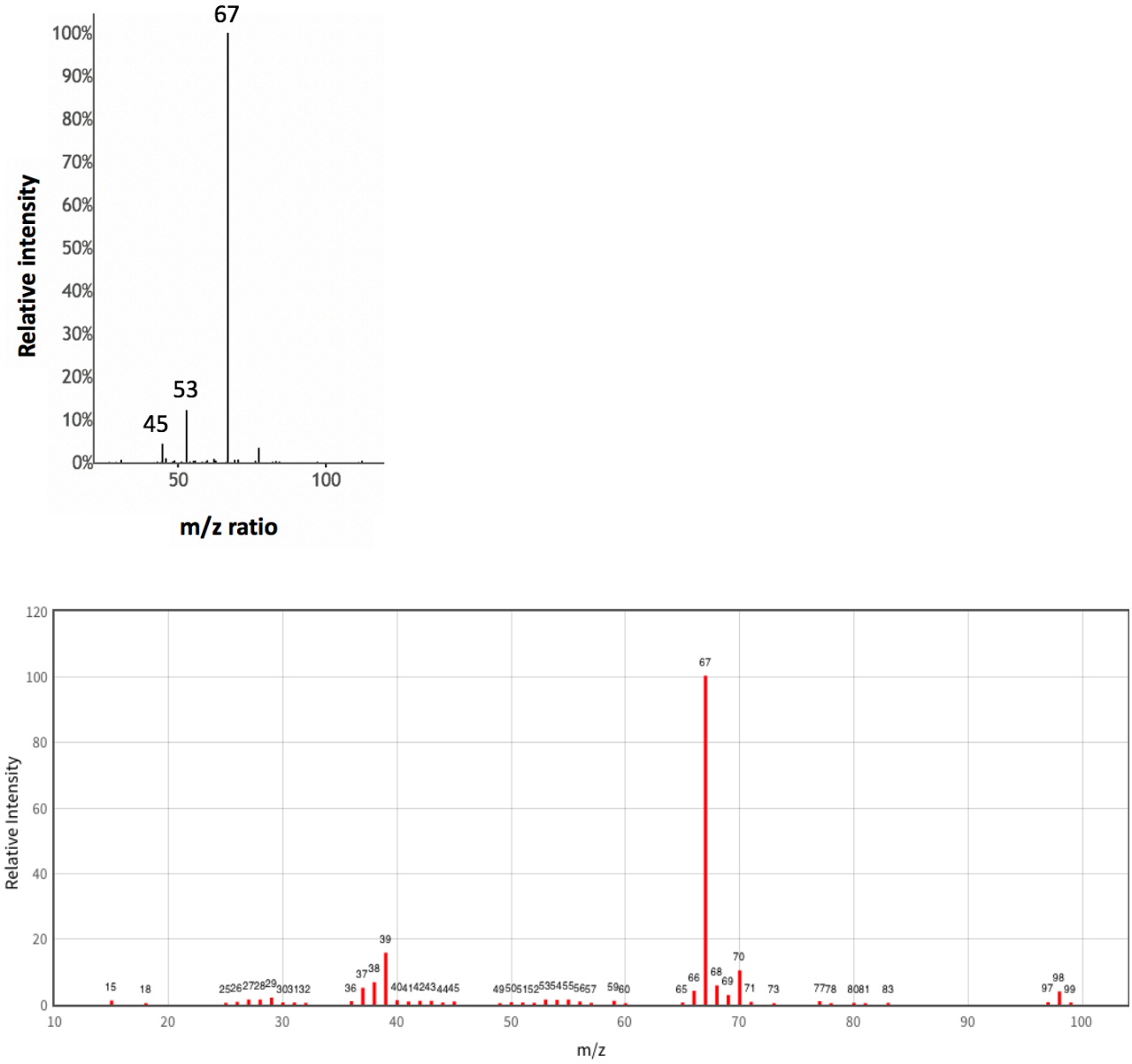


Figure 7-3 Obtained mass spectrum for VOC 12 compared to the mass spectrum of methyl 2-butynoate from the NIST database.

7.11 Appendix 11: Mass spectra and NIST matches for the sulphur compounds

VOC 2 was tentatively identified as dimethyl sulphide, with a good match between the obtained mass spectrum and the NIST database, see Figure 7-4.

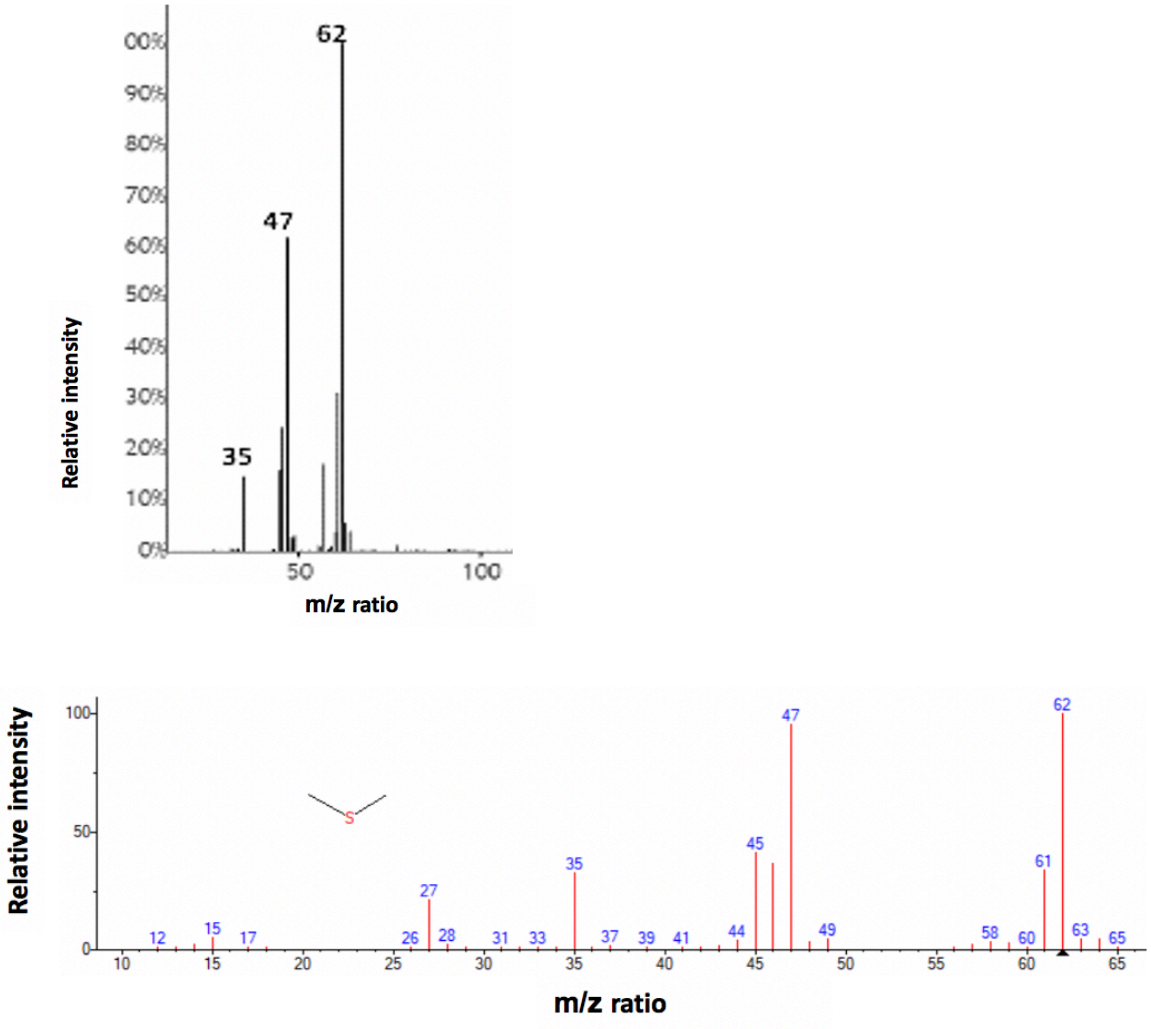


Figure 7-4 Obtained mass spectrum for VOC 2 compared to the mass spectrum of dimethyl sulphide from the NIST database.

7.12 Appendix 12: Mass spectra and NIST matches for the alkanes

VOCs 3 and 11 were identified as two alkanes, likely branched chain alkanes of C14 (see thesis text). As shown in Figure 7-5, the mass spectra for VOC 3 and 11 were very similar, characteristic for alkanes. VOC 15 was tentatively identified as 3-ethyl-hexane, based on the NIST library match, see Figure 7-6.

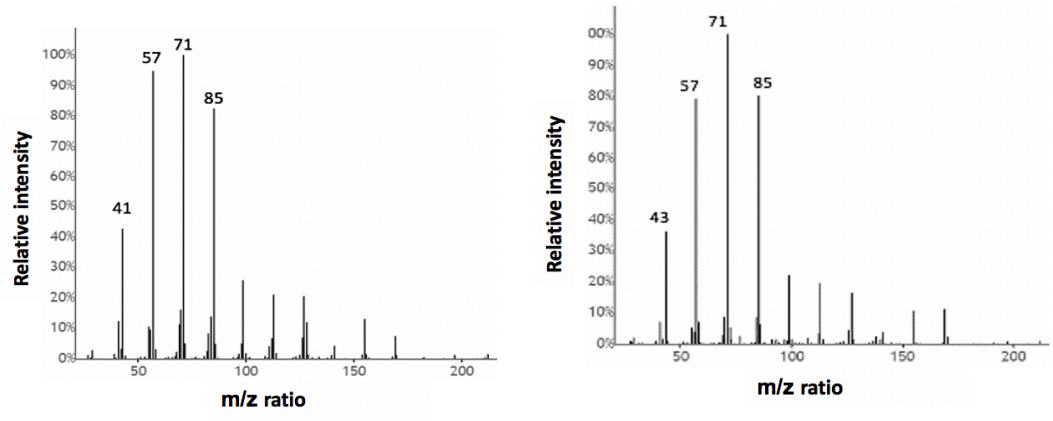


Figure 7-5 Obtained mass spectra for VOC 3 (left) and VOC 11 (right) branched C14 alkanes.

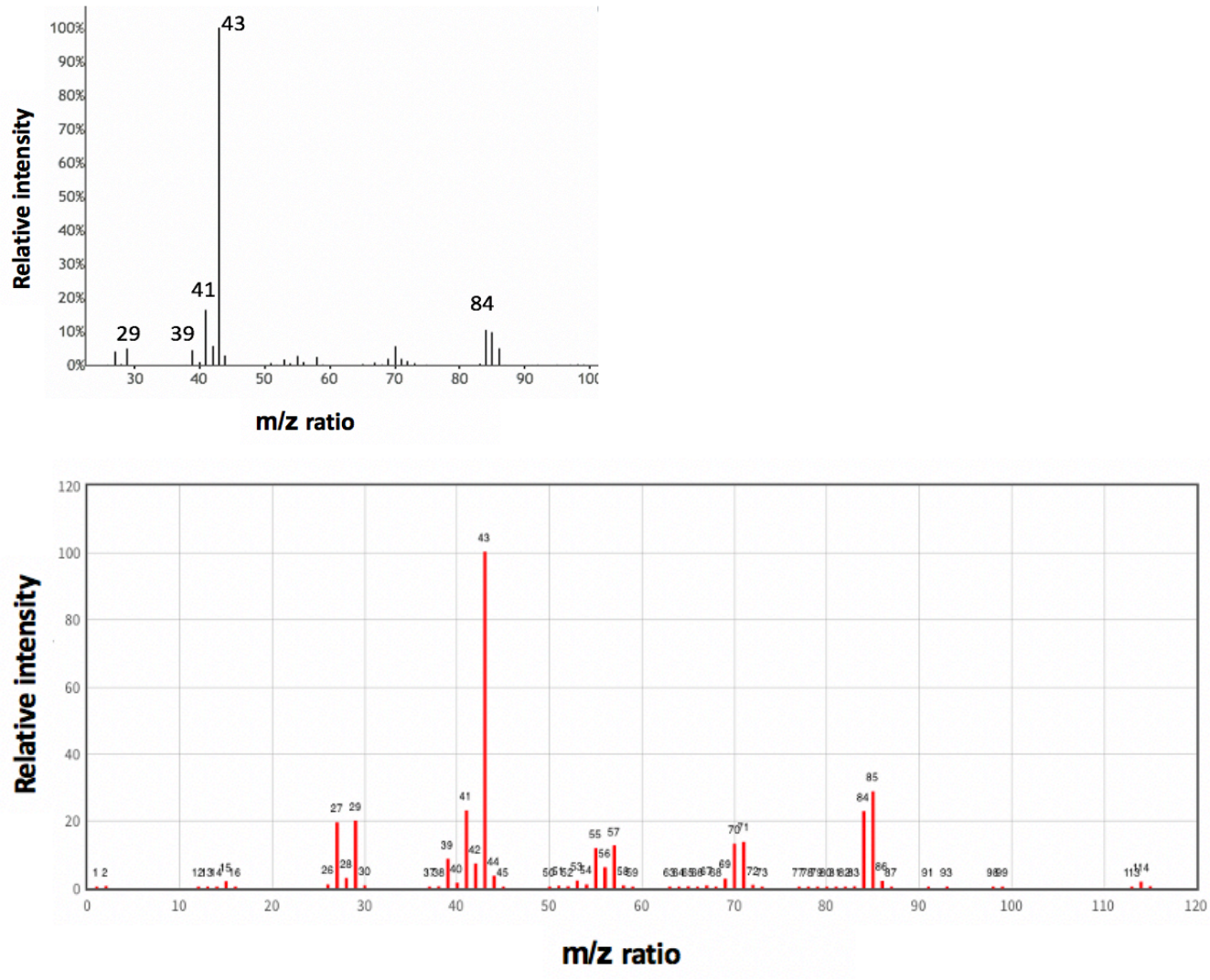


Figure 7-6 Obtained mass spectrum for VOC 15 compared to the mass spectrum of 3-ethyl-hexane from the NIST database.

7.13 Appendix 13: Mass spectra and NIST matches for the alcohols

VOCs 4, 5, 10 and 13 were identified as 1,3-Dioxolane-2-methanol, 2-Phenoxy-ethanol, 2,2,4-Trimethyl-3-pentanol and 1-Undecanol respectively. These all had good matches with corresponding NIST library mass spectra, see Figure 7-7 to Figure 7-10, particularly in the case of VOC 4 and 5, making their tentative identities more confident.

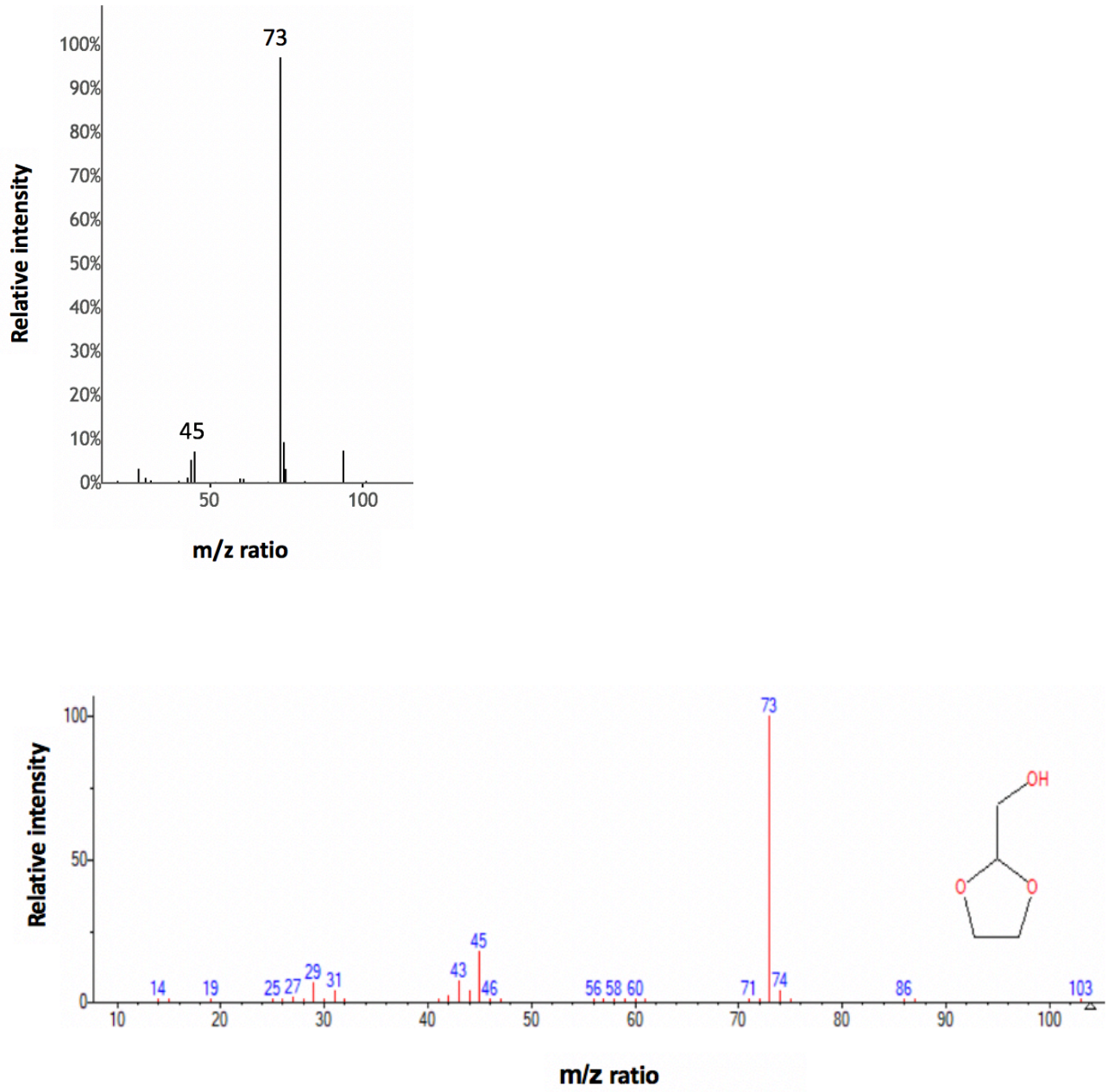


Figure 7-7 Obtained mass spectrum for VOC 4 compared to the mass spectrum of 1,3-dioxolane-2-methanol from the NIST database.

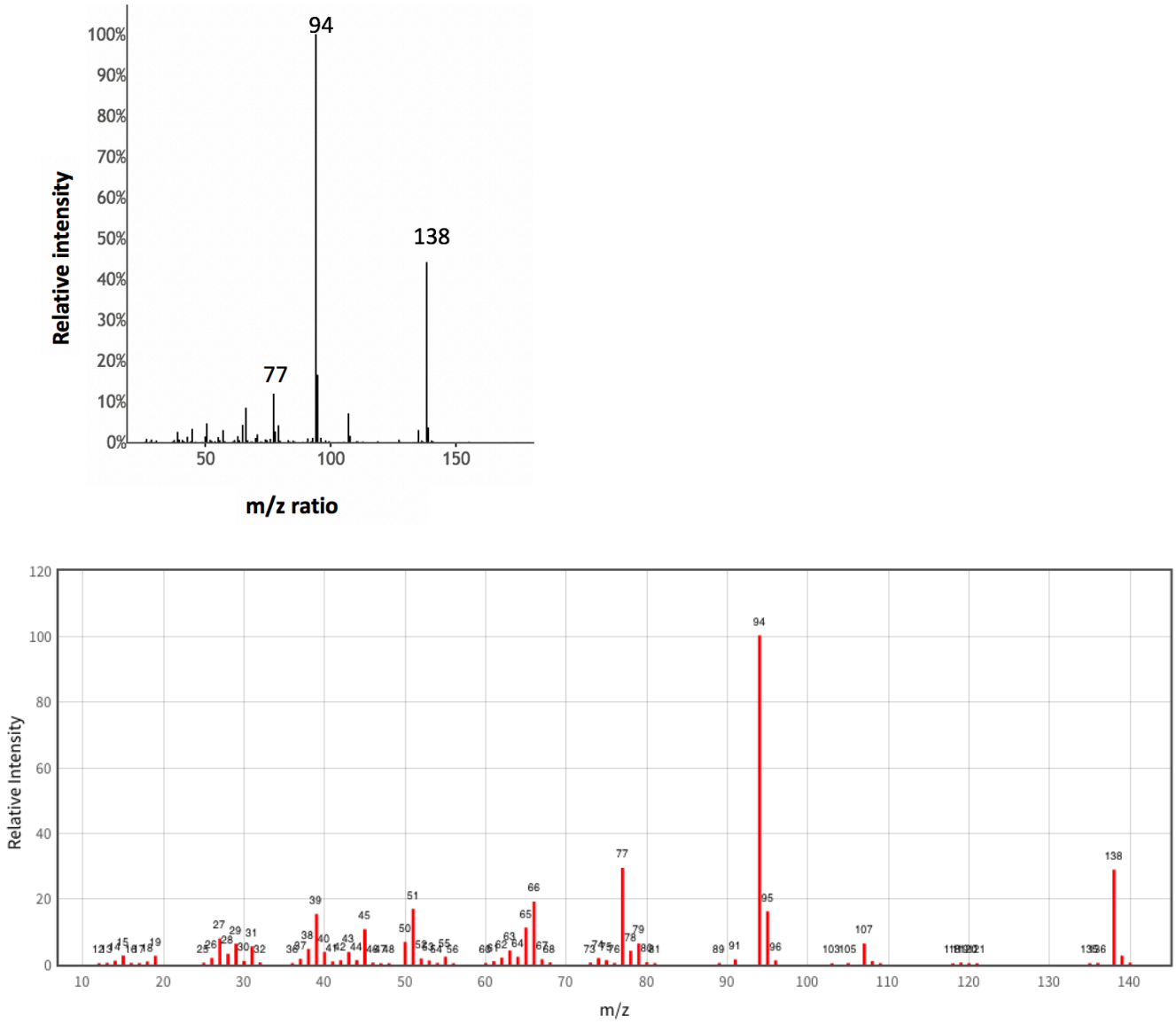


Figure 7-8 Obtained mass spectrum for VOC 5 compared to the mass spectrum of 2-phenoxy-ethanol from the NIST database.

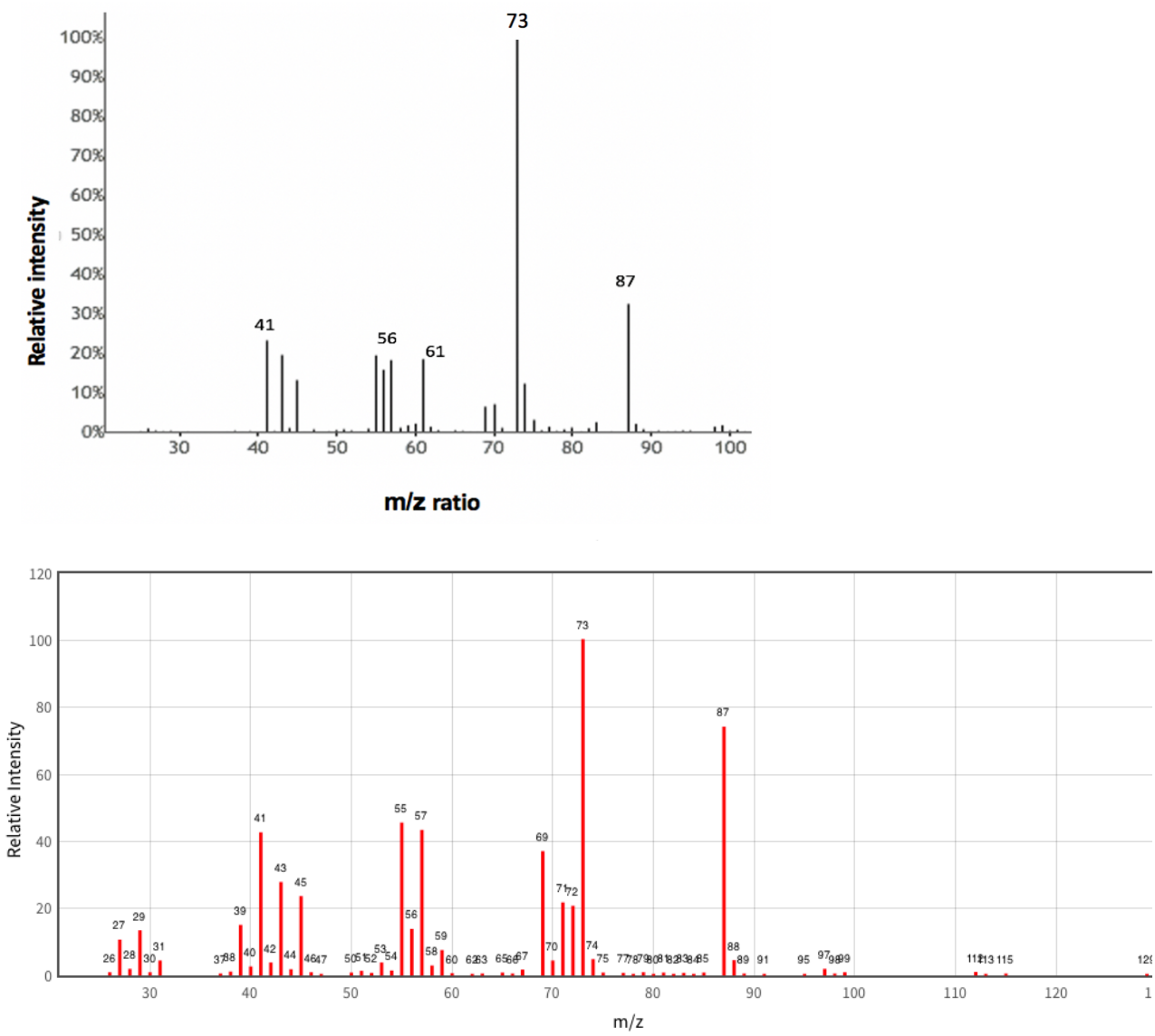
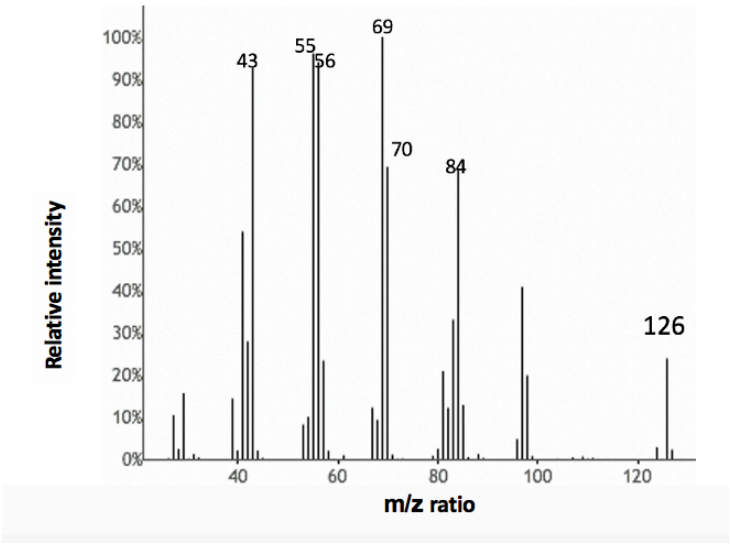


Figure 7-9 Obtained mass spectrum for VOC 10 compared to the mass spectrum of 2,2,4-trimethyl-3-pentanol from the NIST database



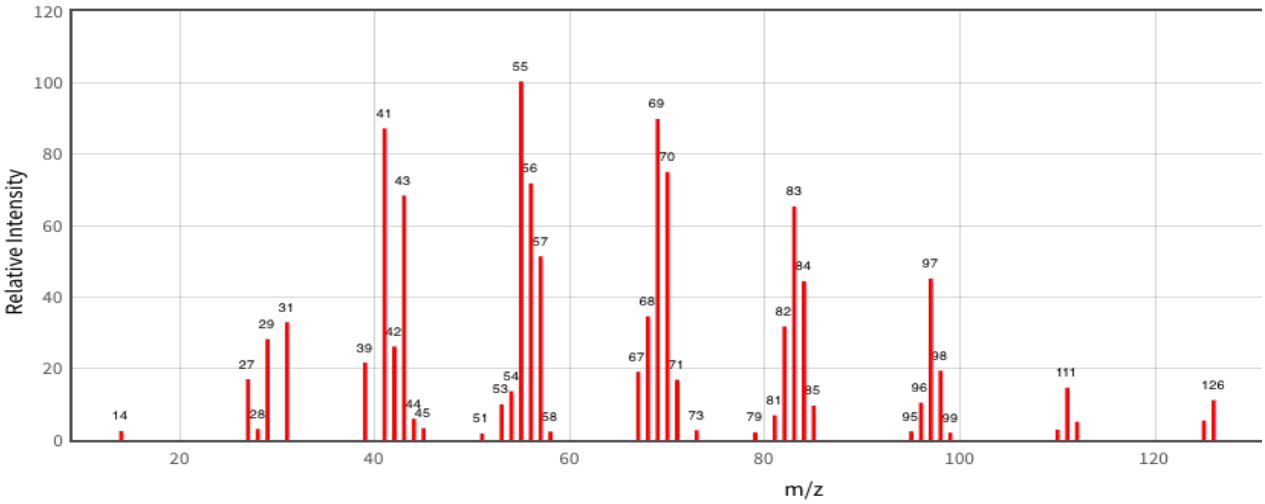


Figure 7-10 Obtained mass spectrum for VOC 13 compared to the mass spectrum of 1-undecanol from the NIST database.

7.14 Appendix 14: Mass spectra and NIST matches for phenol

VOC 14 was identified as phenol on the basis of a good spectral match, indicating confident identification. See Figure 7-11 for the obtained mass spectrum compared to the mass spectrum of phenol from the NIST database.

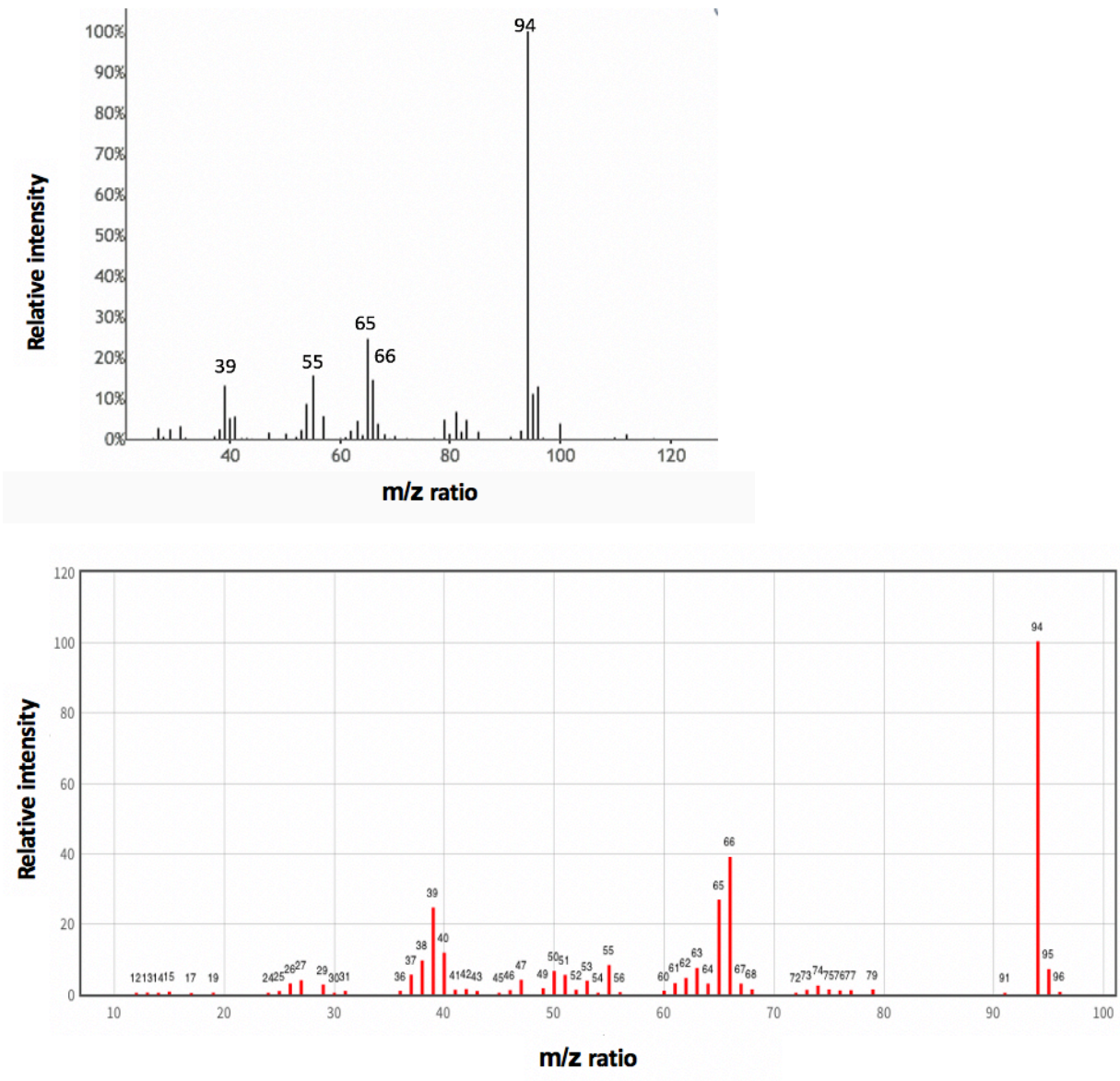
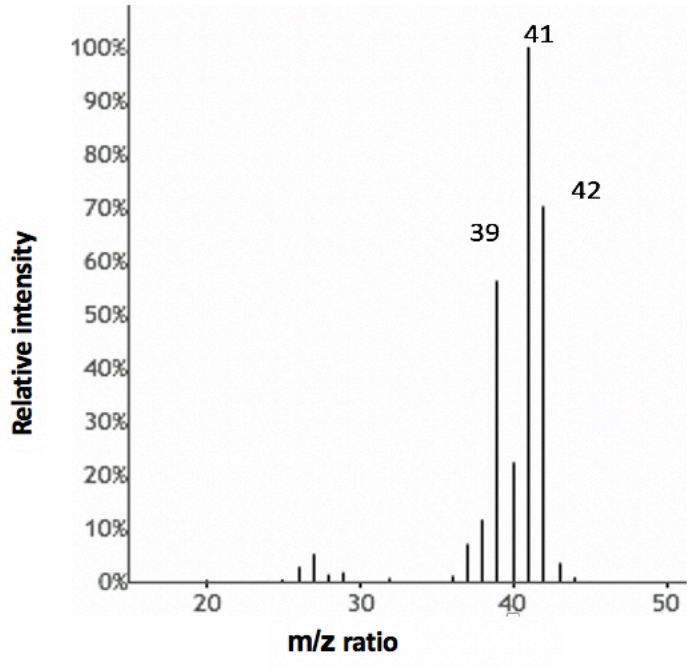


Figure 7-11 Obtained mass spectrum for VOC 14 compared to the mass spectrum of phenol from the NIST database

7.15 Appendix 15: Mass spectra and NIST matches for the non-aromatic cyclic hydrocarbons

VOCs 6 and 7 were identified as cyclopropane and 3,4-dimethyl-1,5-cyclooctadiene. See Figure 7-12 and Figure 7-13 for the obtained mass spectra compared to the comparative mass spectra from the NIST database.



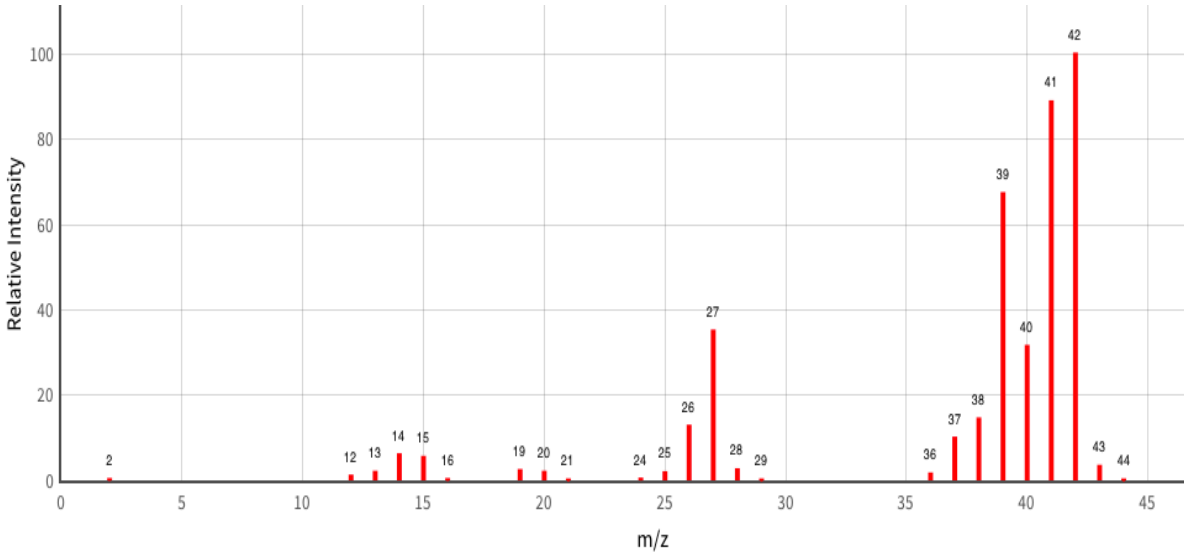


Figure 7-12 Obtained mass spectrum for VOC 6 compared to the mass spectrum of cyclopropane from the NIST database.

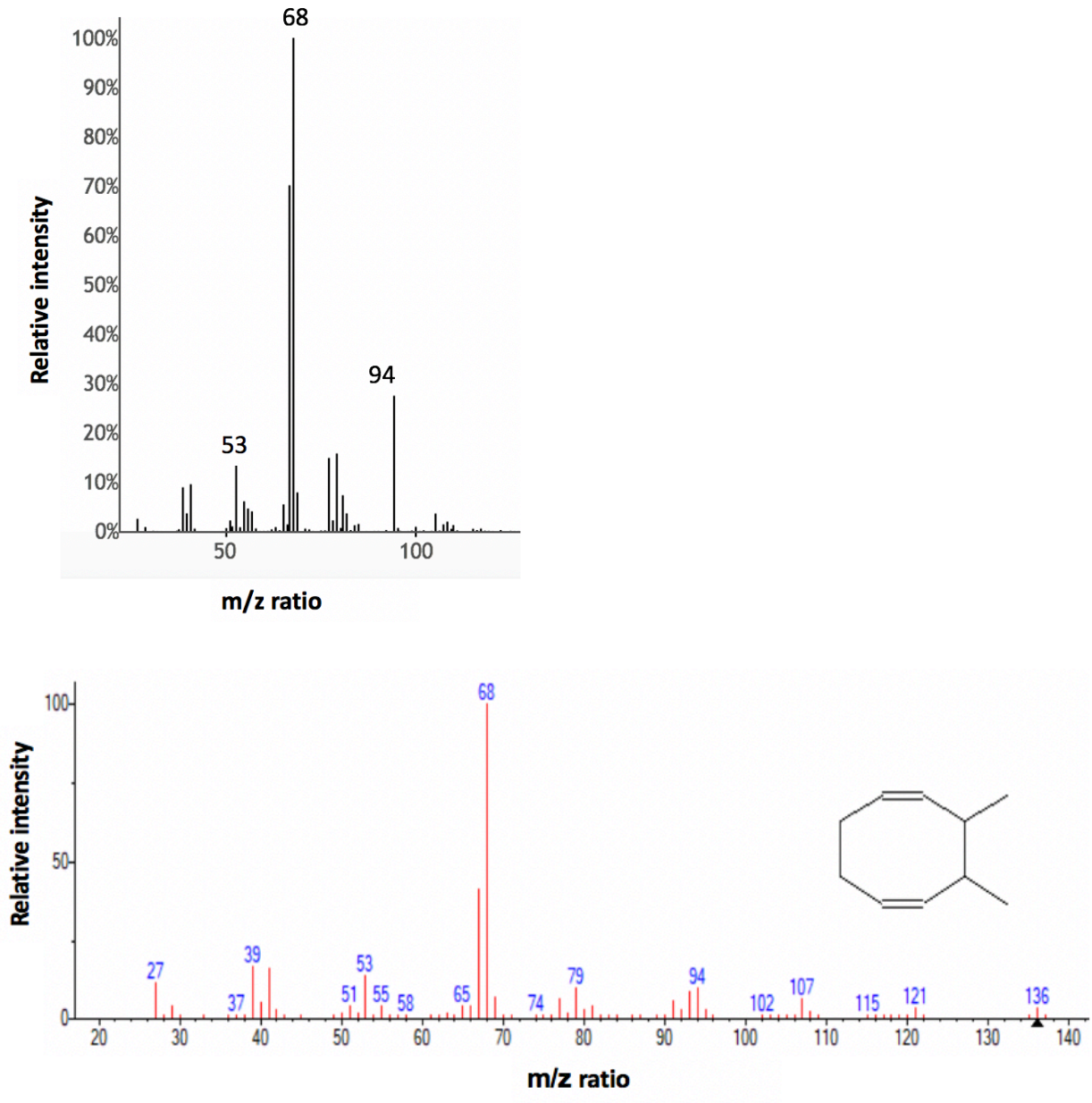


Figure 7-13 Obtained mass spectrum for VOC 7 compared to the mass spectrum of 3,4-dimethyl-1,5-cyclooctadiene from the NIST database.

7.16 **Appendix 16: Names of 26 GP practices participating in the MAGIC breath test study**

Phase-1:

1. Aksyr Medical Practice, NW10 8RY
2. Oxgate Gardens Surgery NW26EA
3. Hillcrest surgery, W3 9RA
4. Dr Jefferies and Partners, Fulham, SW6 6BQ
5. The Law Medical Group practice Willesden NW10 5UY
6. The Law Medical group practice Harrow HA96QQ
7. The Gill medical practice, Feltham TW14 0AB
8. Grove Park Terrace Surgery Chiswick W4
9. The Bush Doctors W12 8PP
10. Twickenham Park Medical Centre, TW13 6HD
11. Buckingham Road Surgery NW10 4RR
12. Fulham Medical Centre, SW6 1BG
13. Acre Surgery, HA6 1TQ
14. Gladstone Medical Centre, NW2 6JH
15. Cuckoo Lane Practice, Hanwell, W7 1DR
16. Wembley Park Medical Centre, Wembley, HA9 8HD

Phase-2:

17. Pimlico Health, SW1V 3EB
18. Lonsdale Medical Centre, NW6 6RR
19. The Good practice, SW10 0LR

7 practices as part of Central London Healthcare(CLH) GP federation:

20. Woodfield Road Medical Centre, W9 3XZ
21. Covent Garden Medical Centre, WC2H 9AA
22. Cavendish Health Centre, W1G 9TG
23. Marylebone Health centre, NW1 5LT
24. Fitzrovia Medical Centre, W1T 6EU
25. Newton Medical Centre, W2 5LT
26. Crawford Street Surgery, W1H 2HJ

7.17 Appendix 17: Summary of National Institute for Health and Care Excellence guidelines for gastrointestinal cancer referral 2016

Available at:

https://www.nice.org.uk/guidance/ng12/chapter/1-Recommendations-organised-by-site-of-cancer#upper-gastrointestinal-tract-cancers

Upper GI cancers

Two week wait (2WW) direct access oesophago-gastro-duodenoscopy (OGD) for:

1) Dysphagia
2) Age >55 years with weight loss AND upper abdominal pain/reflux/dyspepsia

Non urgent direct access OGD:

1) Haematemesis
2) Age >55 years with
   -persistent dyspepsia OR
   -upper abdominal pain WITH anaemia OR
   - raised platelets with nausea/vomiting/weight loss/reflux/dyspepsia/upper abdominal pain OR
   - nausea/vomiting with weight loss/reflux/dyspepsia/upper abdominal pain

2WW CT scan/abdominal ultrasound scan for:

-Abdominal mass
(stomach and gallbladder and liver cancers)

Pancreatic cancer

2WW appointment for:

- Age >40 years with new jaundice

2WW CT scan/ultrasound scan:

- Age >60 years AND weight loss AND diarrhoea/back pain/abdominal pain/nausea/vomiting/constipation/diabetes

Colorectal cancers

2WW appointment for:

1) Age >40 years with weight loss and abdominal pain
2) Age >50 years with rectal bleeding
3) Age >60 years with anaemia/change in bowel habit/positive faecal occult blood test
4) Abdominal mass
5) Age <50 years with rectal bleeding AND abdominal pain/ change in bowel habit /weight loss/anaemia

7.18 **Appendix 18: Poster for GPs**

Imperial College London

Imperial College Healthcare NHS Trust

Breath Testing for Gastrointestinal Disease 2016- 2017

- Please ask all patients with **current/recent** (within 2months) **gastrointestinal symptoms** if they would consider speaking to a research nurse about this study.
- Patients are also eligible with **chronic gastrointestinal problems**, even if controlled on medication.
- Patients must be >18 or <90 years of age, and be able to speak to a research nurse in order to consent to giving a breath sample.

We are testing the feasibility of breath testing for the diagnosis of GI disease.

- **ANY/ALL gastrointestinal symptoms are accepted in this study**. These could include abdominal pain, diarrhoea, constipation, reflux/dyspepsia, dysphagia, change in bowel habit, nausea, vomiting, weight loss, anaemia, GI bleeding, jaundice, or others.

- The study involves giving a breath sample and will take 5-10minutes.
- A research nurse will explain the breath test to the patient and consent them.

Please advise patients to ask at reception to see the clinical research nurses who will be based in the GP practice, or please email: ichc-tr.breathtest@nhs.net with the patient details.

Primary Care Poster v.3.0 16/12/2016. Dr G Woodfield, Prof G Hanna

7.19 Appendix 19: Letter for GPs

Imperial College London

NHS National Institute for Health Research
Clinical Research Network North West London

Division of Surgery
Department of Surgery and Cancer
10th Floor, QEQM Building
St Mary's Hospital
Praed Street, London, W2 1NY

Tel: +44 (0)20 3312 2125
Fax: +44 (0) 020 3312 6309

g.hanna@imperial.ac.uk
www.imperial.ac.uk

Professor George Hanna PhD FRCS
Head of Division of Surgery

14th December 2016

Dear North West London General Practitioners,

Thank you very much for agreeing to participate in the study:

Non-invasive testing for the diagnosis and assessment of gastro-intestinal disease – Primary Care feasibility Study. November 2016 to March 2017.

The study aims to trial the use of a breath test in patients with gastrointestinal symptoms, as a future device for detecting oesophago-gastric, pancreatic and colorectal cancer. This initial trial is a feasibility study of 500 patients. The breath test is not currently a diagnostic tool, as there is no validated "positive" or "negative" result to be gained currently, GPs and patients therefore will not be informed of results. It therefore should not influence patient management or referral pathways in any way.

Recommendation for GPs:

- Please look out for patients who have/have had GI symptoms. Symptoms include **dysphagia, weight loss, abdominal pain, reflux, dyspepsia, nausea, vomiting, diarrhoea, constipation, change in bowel habit, abdominal mass, GI blood loss**, **jaundice** or any other variation of GI symptoms. Patients are eligible if they have had **current/recent symptoms (within 2 months),** OR if they have ever consulted the GP with GI symptoms, where the symptom is **chronic or requires medication to control it**. This includes all patients on frequent anti-reflux medications, laxatives or anti diarrhoeals for example. The symptom does not have to be present on the day of testing.
- Please ask these patients if they would mind talking to a research nurse about performing a breath test as part of our research study. If it is outwith the time that the research nurse is in attendance, please ask them if they would mind being contacted by phone.
- Research nurses will be stationed in practices for 1-2 week blocks. If a research nurse is in your practice that week, please send the patient to see the nurse straight away.
- If it is out of hours or on a week where the research nurse is not present, please email their Name, Phone number and Practice name to ichc-tr.breathtest@nhs.net
- The research nurse will then contact them in a few days time to ask them to come in to perform the breath test.
- Research nurses will give an information leaflet to patients, consent them, document basic medical history and perform the breath test. It should take about 5-10 minutes.
- Practices will be compensated financially for their time in sending patients to see the research nurse for a breath test, and for every patient they highlight via email for the study. (£5 per patient referred)
- Please be aware that neither GPs nor patients will get any feedback/results after their breath test, as both the technology and analysis are still in development. This is not currently a diagnostic tool.

Should you have any questions please don't hesitate to contact me. Thanks again for your cooperation with this study.

Yours sincerely,

Dr Georgia Woodfield, Clinical Research Fellow Imperial College. (g.woodfield@imperial.ac.uk)

Professor George Hanna PhD FRCS, Principle Investigator

7.20 Appendix 20: Participant Information Leaflet- MAGIC study

]

Imperial College London

Department of Surgery & Cancer
Imperial College London
10th Floor, QEQM Building
St Mary's Hospital
Praed Street, London W2 1NY
Tel: +44(0)20 7886 2125
Fax: +44(0)20 7413 0470

Professor George Hanna PhD FRCS
Professor of Surgical Sciences
g.hanna@imperial.ac.uk
www.imperial.ac.uk

PARTICIPANT INFORMATION LEAFLET

Non-invasive diagnostic testing for Gastro-intestinal disease

Principal Investigator:

Professor George Hanna PhD FRCS
Professor of Surgical Sciences / Consultant Surgeon
Imperial College, St Mary's Hospital, London

Co-investigators:

Mr Sheraz Markar MRCS
Clinical Research Fellow
Imperial College London

Mr Tom Wiggins MRCS
Clinical Research Fellow
Imperial College London

Mr. Sacheen Kumar MRCS
Clinical Research Fellow
Imperial College London

Dr. Tanzeela Khalid
Research Scientist
Imperial College London

Ms. Sophie Doran
Clinical Research Fellow
Imperial College London

Dr Georgia Woodfield
Clinical Research Fellow
Imperial College London

Date: 5/7/2017

Dear Sir/Madam

You are being invited to take part in the above titled research study. This information sheet explains the nature of the research being undertaken and what the process involves. Please take your time to read the following information and discuss with others if you wish. Should you require any further information or have any questions, please ask us. Take time to decide whether or not you wish to participate. Thank you for reading this information sheet.

What is the purpose of the study?
We are trying to develop a new way to diagnose cancer of the oesophagus, stomach, colon and rectum in a way that doesn't involve having an endoscopy. To do this we are collecting breath and urine samples. These samples contain metabolites, chemicals produced by cancer cells that can be analysed and may show whether cancer is present. If this analysis is successful we hope to develop a test that can detect cancer without the need for an endoscopy at an early stage when treatment is likely to be more successful.

Why have I been chosen?
You have been chosen for this research project because you have been seen in the GP surgery with gastrointestinal symptoms. We want to see if new non-invasive tests on your breath and urine can also help to diagnose the cause of your symptoms and possibly replace the need for some endoscopies in the future.

Do I have to take part?
Participation in this research study is entirely voluntary. It remains your decision at all times whether or not to take part. If you choose to take part, you will be provided with a copy of this information sheet to keep and requested to sign a consent form. Should you decide at any point that you do not wish to continue with your participation in this research study, you are free to withdraw at any time. A reason for withdrawal need not be given. A decision not to take part or to withdraw at any time shall not affect the standard of care you receive throughout your treatment.

What do I have to do?
You are not required to do anything specific before or after giving your breath sample. You shall undergo your investigations and treatment as planned. The collection of breath and urine will not interfere with your planned treatment nor shall it result in any change to your planned treatment. No additional invasive procedures will be undertaken for the purposes of the research study. We also request your permission to access your medical records for the purpose of the research study only, including blood tests, radiology and pathology results. We will also ask you to fill out an optional anonymous questionnaire about your experience of giving a breath sample. All your medical records shall be handled with strict confidentiality in accordance with the Data Protection Act 1998.

How much of my breath and urine will be taken?
You will be asked to provide a breath sample by breathing normally into a breath sampling device until a defined volume (no more than 2 Litres) has been collected. The device we will be using contains adsorbent tubes that will collect your breath samples. We expect this to take no longer than 5 minutes. You may also be asked to pass a small quantity of urine (20mL) into a container for analysis later. No extra needle pricks or vein punctures are required. The sample shall undergo analysis via our mass spectrometry instruments.

What will happen to the samples?
The machine which we will use to analyse your breath samples will destroy these once we have finished looking at them. Therefore no clinical material will remain. Any breath samples we have not looked at will be discarded. Any surplus urine samples will be stored until the end of the study should we need to re-look at these.

Will I be contacted again in the future?
You shall need to indicate on the consent form whether or not you would like to be contacted again. We shall only contact you to obtain an update on your condition. Whether or not you chose to be contacted shall not impact in any way on the standard of care that you receive; this information may be discussed with you by your Doctor.

What if something goes wrong?
We do not believe that you would be harmed by donating breath or urine samples during this study. Your treatment pathway shall remain the same irrespective of whether or not you choose to participate in this research study. If you wish to complain, or have any concerns about the way you have been approached or treated during the course of this study, the standard National Health Service complaints procedure is available to you.

Will I receive payment for the samples that I donate to the study?
There shall be no payment or remuneration for any sample provided. We shall treat your samples as a gift to the St. Mary's Hospital Upper GI Research Group, London and you would therefore relinquish any interest in the samples provided.

Who is the organising and funding body of the tissue bank and study?
Division of Surgery, Department of Surgery & Cancer, Imperial College London, St. Mary's Hospital, Praed Street, London, W2 1NY.
.
Who has reviewed the ethical considerations of the study?
This study has been reviewed and given favourable ethical opinion for conduct in the NHS by Camden & Islington REC.

Thank you again for taking the time to read this information leaflet. Your participation in this research is most appreciated.

I have some more questions, with whom may I get in contact?

Professor George B Hanna PhD FRCS
Professor of Surgical Sciences / Consultant Upper GI Surgeon
Department of Surgery & Cancer
10th Floor, QEQM Building
St Mary's Hospital, London
Praed street,
London W2 1NY
Tel: 0207 886 2124
Fax: 0207 886 1810
E-mail: g.hanna@imperial.ac.uk

Mr Sheraz Markar
Clinical Research Fellow
Department of Surgery & Cancer
Room 1089
10th Floor, QEQM Building
St Mary's Hospital, London
Praed street,
London W2 1NY
E-mail: s.markar@imperial.ac.uk

Alternatively, you can seek impartial advice from the Patient Advice and Liaison Service (PALS), at PALS, Ground Floor QEQM, St Mary's Hospital, 41 Praed Street, London W2 1NY, Tel: +44(0)2078867777, Fax: +44(0)2078861753

Lastly, the trust R&D provide a third point of contact: Ms Ruth Nicholson AHSC Joint Research Compliance Office, 510B, 5th Floor Lab Block, Charing Cross Hospital, Fulham Palace Road, London, W6 8RF Tel: +44 (0) 203 311 0212 Fax: +44 (0) 203 311 0212 r.nicholson@imperial.ac.uk

7.21 Appendix 21: Consent form – MAGIC study

Imperial College London

Department of Surgery & Cancer
Imperial College London
10th Floor, QEQM Building
St Mary's Hospital
Praed Street, London W2 1NY
Tel: +44(0)20 7886 2125

PARTICIPANT CONSENT FORM

Non-invasive diagnostic testing for Gastro-intestinal disease

Principal Investigator: **Professor George Hanna** PhD FRCS
Professor of Surgical Sciences / Consultant Surgeon
g.hanna@imperial.ac.uk

Please initial box

1. I confirm that I have read and understood the information leaflet for the above study and have had the opportunity to consider the information, ask questions and have had these answered satisfactorily. ☐
2. I understand that my participation is voluntary and that I may withdraw consent to my samples being included in the study at any time without justifying my decision and without affecting my medical care or legal rights. ☐
3. I agree to donate my samples and allow their use in the medical research as described in the Patient Information Leaflet. ☐
4. I understand that sections of any of my medical notes may be looked at by responsible individuals from the research team or from regulatory authorities where it is relevant to my taking part in this research. I give permission for these individuals to have access to my records that are relevant to this research. ☐
5. I understand that my samples will not be used to undertake any genetic tests. ☐
6. I understand that the samples are a gift and that I will not benefit from any intellectual property privilege that result from the use of the samples in research studies. ☐
7. I would like to be contacted if the result of any research carried out on my samples can help me during my treatment course. ☐
8. The samples which I hereby consent to donate are:
   a. Breath ☐
   b. Urine ☐

_____	_____	_____
Name of Patient	Signature	Date
_____	_____	_____
Researcher	Signature	Date

1 copy to be given to the participant, 1 copy in medical notes and original to be filed in trial site file.

Patient consent form Version 8.0/ 10/09/15

7.22 **Appendix 22: Patient acceptability questionnaire**

Non-invasive testing for the diagnosis and assessment of gastro-intestinal disease-
Patient Acceptability Questionnaire

General Practice name…… Study ID…….. Date……..

Please tick the box corresponding to your level of agreement with the following statements:

1.	Yes, today	Yes, in the past 2 months	Yes, over 2 months ago	No, not within the past 5 years	Never
Have you seen a doctor because of stomach/ bowel/abdominal symptoms in the past 5 years?	☐	☐	☐	☐	☐

2.	More than 6 months	Between 2-6 months	Less than 2 months	Less than 1 week	Not applicable
How long did you have your most troubling symptom before seeing a doctor?	☐	☐	☐	☐	☐

3.	Extremely	Quite a bit	Moderately	Slightly	Not at all	Not applicable
How worried were you about your abdominal symptoms when you had them?	☐	☐	☐	☐	☐	☐

4.	Very satisfied	Satisfied	Dissatisfied	Very Dissatisfied	Not applicable
How satisfied were you with the explanation given for how to do the breath test?	☐	☐	☐	☐	☐

5.	Very easy	Easy	Difficult	Very Difficult	Not applicable
How easy was it to do the breath test?	☐	☐	☐	☐	☐

If you found it difficult/very difficult, please explain why……

……

6.	Very comfortable	Comfortable	Uncomfortable	Very uncomfortable	Not applicable
How comfortable were you whilst wearing the face mask?	☐	☐	☐	☐	☐

Patient Questionnaire Version 9.0 5/7/17

7.	Took too long	Acceptable amount of time	Too quick	Not applicable
What did you think about the time it took to give a breath sample?	☐	☐	☐	☐

8.	Very comfortable	Comfortable	Uncomfortable	Very uncomfortable	Not applicable
How did you find the experience of holding the device during the test?	☐	☐	☐	☐	☐

9.	Very comfortable	Comfortable	Uncomfortable	Very uncomfortable	Not applicable
Would you be comfortable to do the breath test again, if recommended by a doctor?	☐	☐	☐	☐	☐

10.	Strongly encourage	Encourage	Discourage	Strongly discourage	Not applicable
Would you encourage family and friends who were offered a breath test to complete it?	☐	☐	☐	☐	☐

11. How could the breath test be improved?

.....……

……

……...……

……..

Please do not hesitate to contact me with further comments or questions:

Dr Georgia Woodfield
Clinical Research Fellow, Imperial College London
g.woodfield@imperial.ac.uk

7.23 **Appendix 23: Questionnaire for GPs**

Non-invasive testing for the diagnosis and assessment of gastro-intestinal disease- Feasibility study

Questionnaire for GPs

Practice name…… Date……..

1. How did you hear about the Breath Test Study? (please circle all that apply)

GP practice meeting GP practice email

Attendance at an NIHR/CRN meeting Colleague

Poster CSO in the GP practice on day of sampling

Other (please specify)……

---

2. What level of information about the study did you receive from the research team regarding: (please circle)

a. Aims of the study	**0**	**1**	**2**	**3**
b. Patient selection by GPs	**0**	**1**	**2**	**3**
c. Recruitment process	**0**	**1**	**2**	**3**
d. General logistics	**0**	**1**	**2**	**3**

0= No information **1= poor/inadequate information**
2= adequate information **3= more than adequate information**

---

3. How could we have improved our methods for disseminating information about the study to GPs/CCGs/practice staff?

---

4. How did you find the Breath Test Study process? (please circle)

a. Asking patients to participate	**0**	**1**	**2**	**3**
b. Sending them to speak to nurse	**0**	**1**	**2**	**3**
c. Answering patient questions	**0**	**1**	**2**	**3**
d. General logistics	**0**	**1**	**2**	**3**

0= Very difficult 1= difficult 2= easy 3= very easy

Please elaborate on any particular barriers ……

GP Questionnaire Version 5.0 16/1/17

5. Would a breath test to detect/rule out cancer be a useful future tool? (please circle)

a. Point of care device with instant results

0 1 2 3

b. Breath test done in same way as a blood test, with results electronically

0 1 2 3

0= Not useful 1= Not sure 2= Useful 3= Very useful

---

6. If you think breath testing could be useful, which patient groups do you think it would particularly benefit?

……

---

7. What do you think is a reasonable cost for GP surgeries to pay for one patient to have a breath test?

……

---

8. How many patients complaining of general gastrointestinal symptoms do you see on average per day as a GP?

……

---

9. Did your request to recruit patients for the study have any impact on the quality of your consultation? (please circle)

0= negative impact
1= no impact
2= minimal impact **0 1 2 3**
3= Positive impact

Please elaborate on any reasons for your answer……..

---

10. Did patients raise questions or concerns when you mentioned the study? If so what were their concerns? .....

……

---

11. How could we have improved the organisation of the study?

……
……

Please do not hesitate to contact me with further comments or questions:
g.woodfield@imperial.ac.uk

GP Questionnaire Version 5.0 16/1/17

7.24 Appendix 24: Breath content quality control for samples run on the PTR-ToF-MS

The same methodology previously described in chapter 3 ("Creation of a sample quality control system (breath content QC) for TD tubes"), was used for breath samples analysed using the PTR-ToF-MS. The same breath samples were used as were used for the development of the GC-MS breath content QC, as only 1 TD tube per sample was ever analysed on the GC-MS, leaving 3 remaining for analysis on the PTR-ToF-MS. 119 breath samples were compared to 138 controls using H_3O^+ ionisation, 109 breath vs 190 controls using NO^+ ionisation, and 122 vs 146 controls using O_2^+ ionisation. The control group were made up of the same sample types as before, blank conditioned TD tubes, TD tubes loaded with room air from different sampling environments, and TD tubes loaded with the previously described standard mix. The thresholds for reference compounds were identified as:

-Acetone >50 ppb for PTR-ToF-MS using H_3O^+ ionisation

-Isoprene >2.5 ppb for PTR-ToF-MS using NO^+ ionisation

-Isoprene >5 ppb for PTR-ToF-MS using O_2^+ ionisation

The development of PTR-ToF-MS thresholds in addition to GC-MS thresholds allowed the same QC processes of tube content QC to be applied to TD tubes regardless of what analytical platform was intended.

If a sample post analysis was found to have acetone or isoprene levels (depending on analysis method used) lower than the level expected from a breath sample, the TD tube was assumed to have an inadequate concentration of breath within it. The breath content data from that tube was therefore excluded for that particular patient's sample.

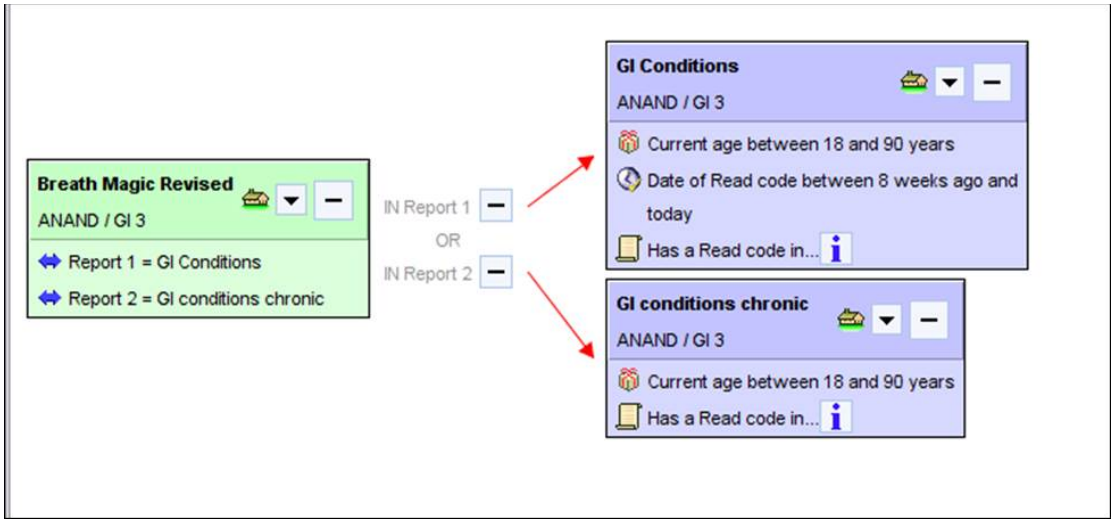
7.25 Appendix 25: Example of GP database searches

The following database searches were performed in order to identify patients for potential Text (SMS) recruitment during Phase-2 of Breath MAGIC, as part of the hub and spoke model of sampling. Of note, this search example is the most complex and thorough search done within the Breath MAGIC study. In Phase-1, local GP practices did simple local database searches performed by local GP receptionists or GPs themselves based on specific symptoms or medication use. This is because only small numbers of patients were needed per practice in Phase-1. The aim of this more complex search was to reach as many eligible patients as possible from 7 GP practices in the Central London Healthcare (CLH) GP federation for General Practices in Westminster, to be breath-tested at the central hub (Marylebone Health Centre). The database search was performed by central CLH administrative staff at CCG level (Ahmed Hosny and Anand Bhundia- GP Network Support Officers CLH), neither of whom worked directly in the participating practices. The large searches of CLH records for the hub and spoke sampling strategy was done as follows:

Search 1

Gastrointestinal symptoms recorded in the past eight weeks (including those coded as chronic) AND age 18-90 years inclusive

Search 1 therefore picked up patients who fulfilled the 1st and 2nd inclusion criteria for the study (gastrointestinal symptoms today or within last 8 weeks).



NB: gastrointestinal symptoms are represented by “i” in the diagram above. Symptoms (with database read codes) included:

Indigestion (1954.)
Abdominal pain (1969.)
Altered bowel function (19EA.)
Diarrhoea (19F2.)
Viral gastroenteritis (A07y0)
Gastro-oesophageal reflux disease with ulceration (J1020)
[D]Dysphagia (R072.)
[D]Change in bowel habit (R078.)
[D]Abdominal pain (R090.)
Gastric reflux (Ua1kQ)
Gastro-oesophageal reflux disease (X3003)
Gastritis (X301N)
Gastroenteritis (X30BN)
Nausea (X75qw)
Jaundice (X769z)
Weight loss (X76CA)
Flatulent dyspepsia (X76d5)
Campylobacter gastrointestinal tract infection (XE0Ql)
Gastro-oesophageal reflux disease with oesophagitis (XE0aL)
Gastro-oesophageal reflux disease without oesophagitis (XE0aO)
Irritable bowel syndrome (XE0as)
Biliary tract disorders NOS (XE0dR)
Constipation (XE0rD)
[D]Abdominal mass (XE2nV)
Dysphagia (XM08J)
Abdominal mass (XM097)
Bacterial gastroenteritis (XM0pJ)
Nausea and vomiting (Xa1pJ)
Moderate gastric reflux (Xa7Ta)
Minimal gastric reflux (Xa7Tb)
Gastric aspirate containing blood (Xa7Tj)

Chronic conditions:
Chronic gastric ulcer (J111.)
Chronic gastrojejunal ulcer (J141.)
Chronic gastritis (J151.)
Chronic constipation with overflow (J5201)
Chronic nonspecific abdominal pain (X3062)
Chronic constipation (X30Bl)
Chronic diarrhoea (X30Bn)

Search 2

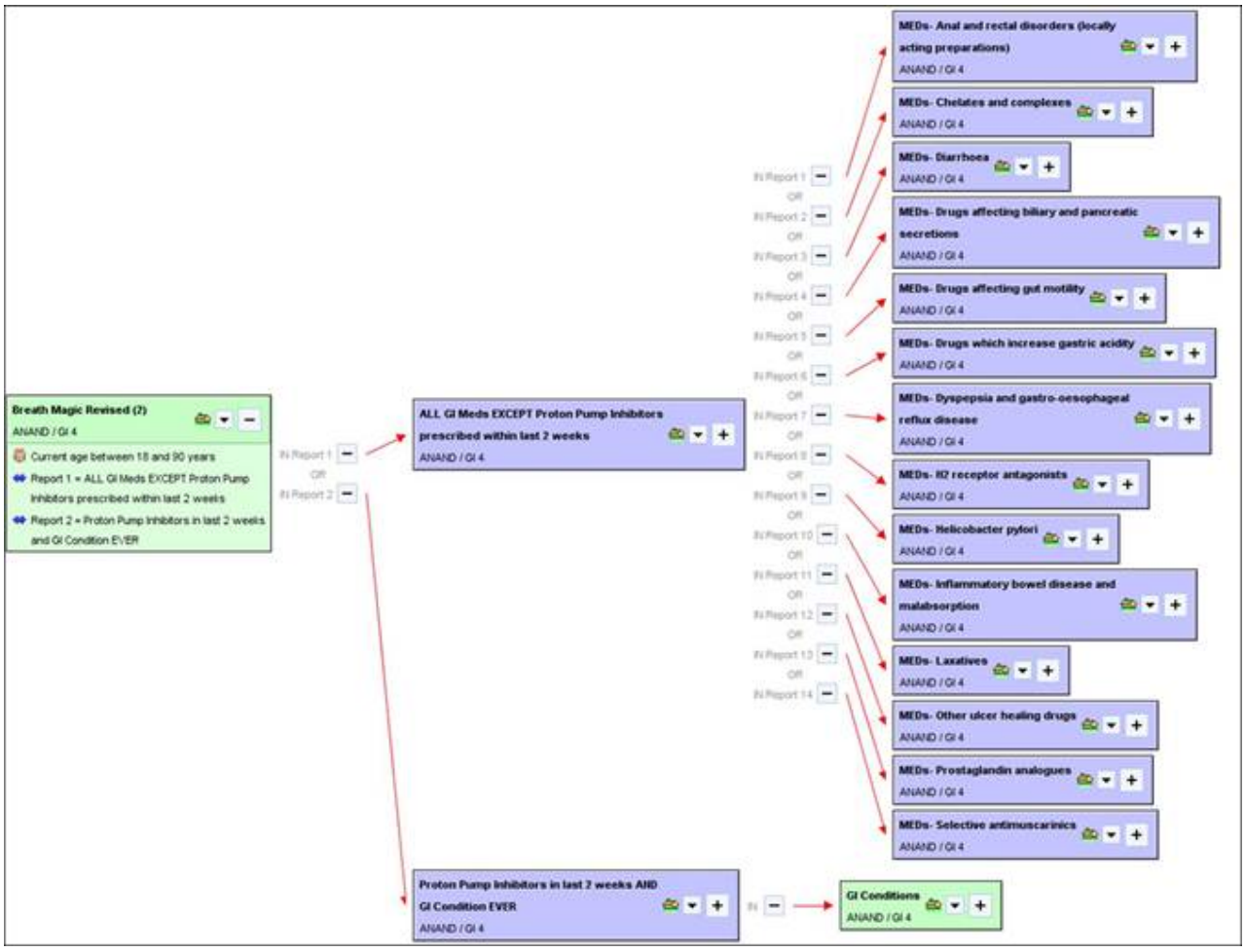
Currently on gastrointestinal medications except proton pump inhibitors (PPIs) prescribed within last two weeks AND age 18-90 years inclusive.

OR

Patients prescribed PPIs (within last two weeks) AND a recorded gastrointestinal condition at any point in their records.

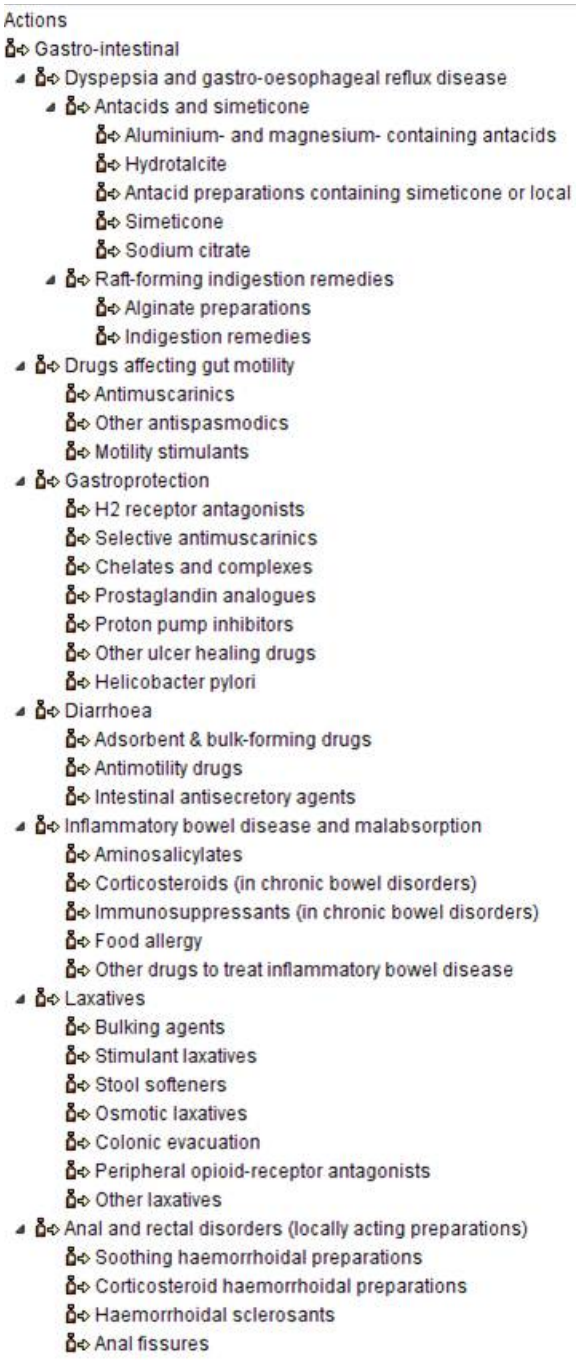
Search 2 therefore picked up patients who fulfilled the 3rd inclusion criteria for the study (chronic gastrointestinal condition controlled on medication).

The caveat with PPIs was that we did not want patients who were on PPIs for non-gastrointestinal reasons; e.g. patients who were on prophylactic PPIs because they were taking steroids. For this reason, the search started with "All GI meds EXCEPT PPIs prescribed in last two weeks" AND "PPIs in last two weeks AND GI condition EVER".



Gastrointestinal medications that were included as part of this search are encoded by the headings on the right of the diagram above.

Included categories and subcategories of gastrointestinal medications were:



Results of searches 1 and 2, hub and spoke model

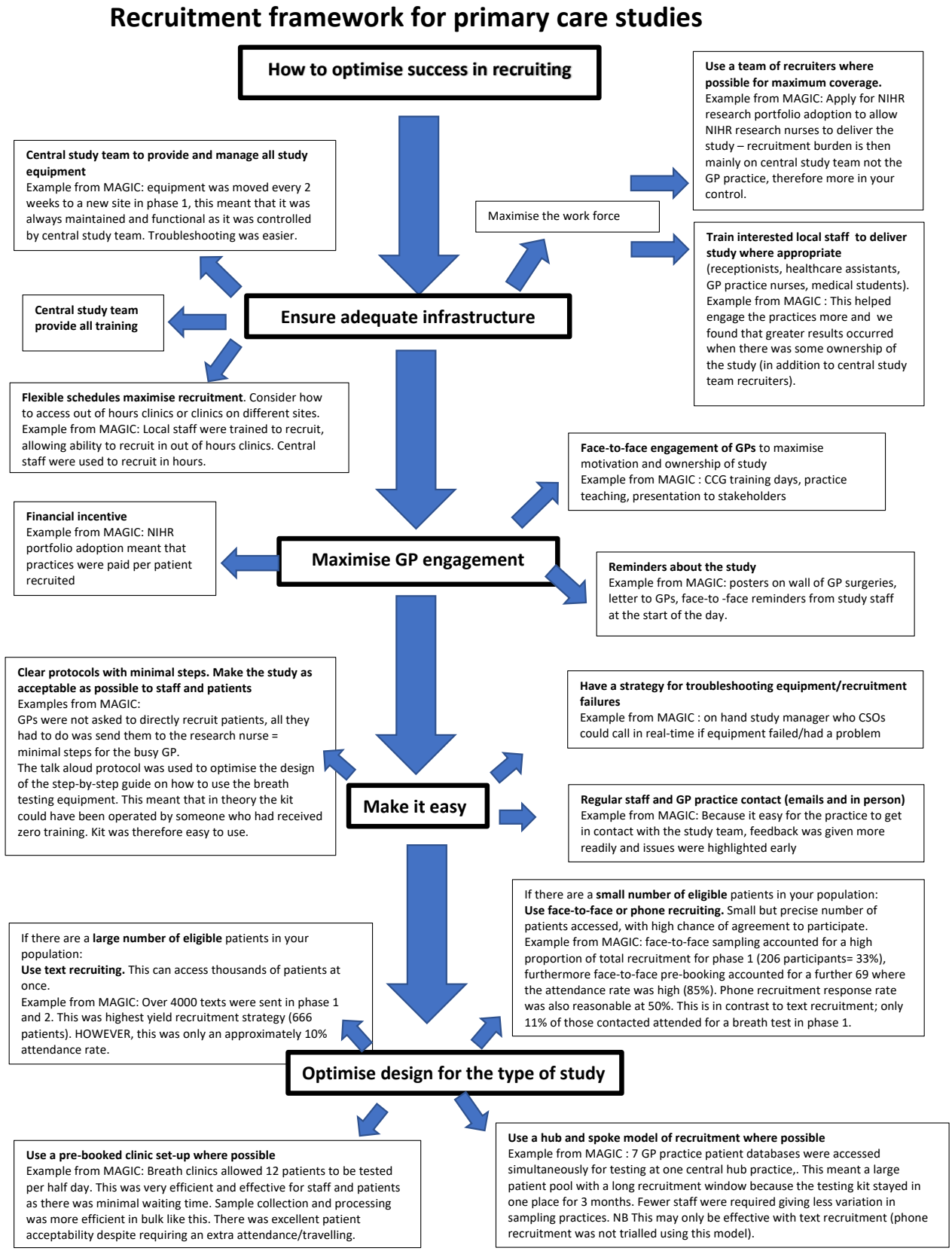
The search identified 1293 potential patients, with a final yield of 103 successful breath tests. See table 7-1 for breakdown of responses. In summary, 22% of people of

replied “yes” to the invitation text. Of these, 55% booked a breath testing appointment. Of those who booked, there was 70% successful breath test rate.

Table 7-1 Breakdown of texts sent and responses

TOTALS	
7	**Practices**
1293	**Search**
176	Excluded before texting
1117	Texted
281	Replied YES
267	**Called**
6	No Answer
55	Voicemail (6 successfully called back)
35	Not Eligible
29	Declined
148	**Booked**
42	DNA
3	Face to Face decline
103	**Attended, consented and performed breath test**

7.26 Appendix 26: Optimising recruitment in primary care studies, lessons learnt from the MAGIC breath testing study.



8 REFERENCES AND BIBLIOGRAPHY

1. Cancer research UK. Bowel cancer statistics. 2017 [updated 11/3/20. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer#heading-Zero.
2. Office for National Statistics. Cancer Registration Statistics, England. 2017 [updated 26/4/19. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/cancerregistrationstatisticscancerregistrationstatisticsengland.
3. Office for National Statistics. Cancer survival by stage at diagnosis for England. 2019 [updated 2/4/20. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/datasets/cancersurvivalratescancersurvivalinenglandadultsdiagnosed.
4. De Angelis R, Sant M, Coleman MP, Francisci S, Baili P, Pierannunzio D, et al. Cancer survival in Europe 1999-2007 by country and age: results of EUROCARE-5- a population-based study. The Lancet Oncology. 2014;15(1):23-34.
5. Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, et al. Genetic alterations during colorectal-tumor development. The New England journal of medicine. 1988;319(9):525-32.
6. Schofield PF JD. ABC of colorectal diseases. Colorectal neoplasia—I: Benign colonic tumours. BMJ (Clinical research ed). 1992;394:1498-500.
7. Rutter MD, East J, Rees CJ, Cripps N, Docherty J, Dolwani S, et al. British Society of Gastroenterology/Association of Coloproctology of Great Britain and Ireland/Public Health England post-polypectomy and post-colorectal cancer resection surveillance guidelines. Gut. 2020;69(2):201-23.
8. East JE, Atkin WS, Bateman AC, Clark SK, Dolwani S, Ket SN, et al. British Society of Gastroenterology position statement on serrated polyps in the colon and rectum. Gut. 2017;66(7):1181-96.
9. Grady WM, Markowitz SD. The molecular pathogenesis of colorectal cancer and its potential application to colorectal cancer screening. Digestive diseases and sciences. 2015;60(3):762-72.
10. Cancer Research UK. Bowel cancer survival statistics - by stage. 2018 [Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/bowel-cancer/survival#heading-Three.
11. Richards MA. The size of the prize for earlier diagnosis of cancer in England. Br J Cancer. 2009;101(S2):S125-S9.
12. National Bowel Cancer Audit. Annual Report 2020. UK. [updated December 2020. Available from: https://www.nboca.org.uk/content/uploads/2020/12/NBOCA-2020-Annual-Report.pdf.
13. Atkin W, Wooldrage K, Parkin DM, Kralj-Hans I, MacRae E, Shah U, et al. Long term effects of once-only flexible sigmoidoscopy screening after 17 years of follow-up: the UK Flexible Sigmoidoscopy Screening randomised controlled trial. The Lancet. 2017;389(10076):1299-311.
14. Pox CP, Altenhofen L, Brenner H, Theilmeier A, Von Stillfried D, Schmiegel W. Efficacy of a nationwide screening colonoscopy program for colorectal cancer. Gastroenterology. 2012;142(7):1460-7.e2.
15. Bowel cancer UK. Improving capacity, saving lives: endoscopy in the UK. London: Bowel Cancer UK. 2012:10.

16. Lee TJ, Siau K, Esmaily S, Docherty J, Stebbing J, Brookes MJ, et al. Development of a national automated endoscopy database: The United Kingdom National Endoscopy Database (NED). United European Gastroenterol J. 2019;7(6):798-806.
17. Vega P, Valentín F, Cubiella J. Colorectal cancer diagnosis: Pitfalls and opportunities. World J Gastrointest Oncol. 2015;7(12):422-33.
18. Rutter CM, Johnson E, Miglioretti DL, Mandelson MT, Inadomi J, Buist DS. Adverse events after screening and follow-up colonoscopy. Cancer Causes Control. 2012;23(2):289-96.
19. 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.
20. National Institute of Clinical Excellence. Suspected cancer: recognition and referral. NICE guideline [NG12]. Lower gastrointestinal tract cancers. UK 2015 [updated July 2017] [Available from: https://www.nice.org.uk/guidance/ng12/chapter/1-Recommendations-organised-by-site-of-cancer#lower-gastrointestinal-tract-cancers.
21. Cross AJ, Wooldrage K, Robbins EC, Pack K, Brown JP, Hamilton W, et al. Whole-colon investigation vs. flexible sigmoidoscopy for suspected colorectal cancer based on presenting symptoms and signs: a multicentre cohort study. British journal of cancer. 2019;120(2):154-64.
22. National Institute of Clinical Excellence. Quantitative faecal immunochemical tests to guide referral for colorectal cancer in primary care. Diagnostics guidance [DG30]. London 2017 [Available from: https://www.nice.org.uk/guidance/dg30/chapter/2-Clinical-need-and-practice.
23. Farrugia A, Widlak M, Evans C, Smith SC, Arasaradnam R. Faecal immunochemical testing (FIT) in symptomatic patients: what are we missing? Frontline Gastroenterology. 2020;11(1):28-33.
24. Westwood M, Corro Ramos I, Lang S, Luyendijk M, Zaim R, Stirk L, et al. Faecal immunochemical tests to triage patients with lower abdominal symptoms for suspected colorectal cancer referrals in primary care: a systematic review and cost-effectiveness analysis. Health Technol Assess. 2017;21(33):1-234.
25. Westwood M, Lang S, Armstrong N, van Turenhout S, Cubiella J, Stirk L, et al. Faecal immunochemical tests (FIT) can help to rule out colorectal cancer in patients presenting in primary care with lower abdominal symptoms: a systematic review conducted to inform new NICE DG30 diagnostic guidance. BMC Med. 2017;15(1):189.
26. Imperiale TF, Ransohoff DF, Itzkowitz SH, Levin TR, Lavin P, Lidgard GP, et al. Multitarget Stool DNA Testing for Colorectal-Cancer Screening. New England Journal of Medicine. 2014;370(14):1287-97.
27. Hewitson P, Glasziou P, Watson E, Towler B, Irwig L. Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (hemoccult): an update. Am J Gastroenterol. 2008;103(6):1541-9.
28. Mandel JS, Bond JH, Church TR, Snover DC, Bradley GM, Schuman LM, et al. Reducing mortality from colorectal cancer by screening for fecal occult blood. Minnesota Colon Cancer Control Study. The New England journal of medicine. 1993;328(19):1365-71.
29. Hardcastle JD, Chamberlain JO, Robinson MH, Moss SM, Amar SS, Balfour TW, et al. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. Lancet. 1996;348(9040):1472-7.

30. Kronborg O, Fenger C, Olsen J, Jørgensen OD, Søndergaard O. Randomised study of screening for colorectal cancer with faecal-occult-blood test. Lancet. 1996;348(9040):1467-71.
31. Lieberman DA. Clinical practice. Screening for colorectal cancer. The New England journal of medicine. 2009;361(12):1179-87.
32. Ahlquist DA, Shuber AP. Stool screening for colorectal cancer: evolution from occult blood to molecular markers. Clin Chim Acta. 2002;315(1-2):157-68.
33. Ahlquist DA, Wieand HS, Moertel CG, McGill DB, Loprinzi CL, O'Connell MJ, et al. Accuracy of fecal occult blood screening for colorectal neoplasia. A prospective study using Hemoccult and HemoQuant tests. The Journal of the American Medical Association. 1993;269(10):1262-7.
34. Allison JE, Tekawa IS, Ransom LJ, Adrain AL. A Comparison of Fecal Occult-Blood Tests for Colorectal-Cancer Screening. New England Journal of Medicine. 1996;334(3):155-60.
35. Imperiale TF, Ransohoff DF, Itzkowitz SH, Turnbull BA, Ross ME. Fecal DNA versus Fecal Occult Blood for Colorectal-Cancer Screening in an Average-Risk Population. New England Journal of Medicine. 2004;351(26):2704-14.
36. von Karsa L, Patnick J, Segnan N, Atkin W, Halloran S, Lansdorp-Vogelaar I, et al. European guidelines for quality assurance in colorectal cancer screening and diagnosis: overview and introduction to the full supplement publication. Endoscopy. 2013;45(1):51-9.
37. van Rossum LG, van Rijn AF, Laheij RJ, van Oijen MG, Fockens P, van Krieken HH, et al. Random comparison of guaiac and immunochemical fecal occult blood tests for colorectal cancer in a screening population. Gastroenterology. 2008;135(1):82-90.
38. Benson VS, Atkin WS, Green J, Nadel MR, Patnick J, Smith RA, et al. Toward standardizing and reporting colorectal cancer screening indicators on an international level: The international colorectal cancer screening network. International Journal of Cancer. 2012;130(12):2961-73.
39. Stonestreet J, Chandrapalan S, Woolley D, Uthman U, Arasaradnam RP. Systematic review and meta-analysis : diagnostic accuracy of faecal immunochemical testing for haemoglobin (FIT) in detecting colorectal cancer for both symptomatic and screening population. Acta Gastroenterol Belg. 2019;82(2):291-9.
40. Fraser CG. Faecal occult blood tests--eliminate, enhance or update? Annals of clinical biochemistry. 2008;45(Pt 2):117-21.
41. Cancer research UK. Faecal Immunochemical Test (FIT). UK 2020 [updated 21/3/2020. Available from: https://www.cancerresearchuk.org/health-professional/screening/bowel-screening-evidence-and-resources/faecal-immunochemical-test-fit#FIT2.
42. Imperiale TF, Gruber RN, Stump TE, Emmett TW, Monahan PO. Performance Characteristics of Fecal Immunochemical Tests for Colorectal Cancer and Advanced Adenomatous Polyps. Annals of internal medicine. 2019;170(5):319-29.
43. Peacock O, Clayton S, Atkinson F, Tierney GM, Lund JN. 'Be Clear on Cancer': the impact of the UK National Bowel Cancer Awareness Campaign. Colorectal Disease. 2013;15(8):963-7.
44. Thorne K, Hutchings HA, Elwyn G. The effects of the Two-Week Rule on NHS colorectal cancer diagnostic services: A systematic literature review. BMC Health Services Research. 2006;6(1):43.

45. Elliss-Brookes L, McPhail S, Ives A, Greenslade M, Shelton J, Hiom S, et al. Routes to diagnosis for cancer - determining the patient journey using multiple routine data sets. Br J Cancer. 2012;107(8):1220-6.
46. Sheringham JR, Georghiou T, Chitnis XA, Bardsley M. Comparing primary and secondary health-care use between diagnostic routes before a colorectal cancer diagnosis: Cohort study using linked data. Br J Cancer. 2014;111(8):1490-9.
47. Public Health England. National Cancer Intelligence Network: Routes to diagnosis 2006-2016. London2016 [updated April 2019. Available from: http://www.ncin.org.uk/publications/routes_to_diagnosis.
48. Logan RF, Patnick J, Nickerson C, Coleman L, Rutter MD, von Wagner C, et al. Outcomes of the Bowel Cancer Screening Programme (BCSP) in England after the first 1 million tests. Gut. 2012;61(10):1439-46.
49. Moss SM, Hardcastle JD, Coleman DA, Robinson MH, Rodrigues VC. Interval cancers in a randomized controlled trial of screening for colorectal cancer using a faecal occult blood test. International journal of epidemiology. 1999;28(3):386-90.
50. Gyrd-Hansen D, Sogaard J, Kronborg O. Analysis of screening data: colorectal cancer. International journal of epidemiology. 1997;26(6):1172-81.
51. Church TR, Ederer F, Mandel JS. Fecal occult blood screening in the Minnesota study: sensitivity of the screening test. Journal of the National Cancer Institute. 1997;89(19):1440-8.
52. Lee TJ, Clifford GM, Rajasekhar P, Rutter MD, Kometa S, Ritchie MC, et al. High yield of colorectal neoplasia detected by colonoscopy following a positive faecal occult blood test in the NHS Bowel Cancer Screening Programme. Journal of medical screening. 2011;18(2):82-6.
53. Palmer CK, Thomas MC, von Wagner C, Raine R. Reasons for non-uptake and subsequent participation in the NHS Bowel Cancer Screening Programme: a qualitative study. British Journal of Cancer. 2014;110(7):1705-11.
54. Moss S, Mathews C, Day TJ, Smith S, Seaman HE, Snowball J, et al. Increased uptake and improved outcomes of bowel cancer screening with a faecal immunochemical test: results from a pilot study within the national screening programme in England. Gut. 2017;66(9):1631-44.
55. NHS England. Report of the independent reveiw of adult screening programmes in England 2019 [Available from: https://www.england.nhs.uk/wp-content/uploads/2019/02/report-of-the-independent-review-of-adult-screening-programme-in-england.pdf.
56. Fraser CG, Mathew CM, Mowat NAG, Wilson JA, Carey FA, Steele RJC. Evaluation of a card collection-based faecal immunochemical test in screening for colorectal cancer using a two-tier reflex approach. Gut. 2007;56(10):1415-8.
57. Clarke P, Jack F, Carey FA, Steele RJC. Medications with anticoagulant properties increase the likelihood of a negative colonoscopy in faecal occult blood test population screening. Colorectal Disease. 2006;8(5):389-92.
58. National Institute of Clinical Excellence. Quantitative faecal immunochemical tests to guide referral for colorectal cancer in primary care
Diagnostics guidance [DG30]. Committee discussion.: London; 2017 [Available from: https://www.nice.org.uk/guidance/dg30/chapter/5-Committee-discussion.
59. Campos-da-Paz M, Dórea JG, Galdino AS, Lacava ZGM, de Fatima Menezes Almeida Santos M. Carcinoembryonic Antigen (CEA) and Hepatic Metastasis in Colorectal Cancer:

Update on Biomarker for Clinical and Biotechnological Approaches. Recent Pat Biotechnol. 2018;12(4):269-79.
60. Gonzalez-Pons M, Cruz-Correa M. Colorectal Cancer Biomarkers: Where Are We Now? Biomed Res Int. 2015;2015:149014.
61. Shah R, Jones E, Vidart V, Kuppen PJ, Conti JA, Francis NK. Biomarkers for early detection of colorectal cancer and polyps: systematic review. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2014;23(9):1712-28.
62. Peltier J, Roperch JP, Audebert S, Borg JP, Camoin L. Quantitative proteomic analysis exploring progression of colorectal cancer: Modulation of the serpin family. J Proteomics. 2016;148:139-48.
63. Ghazanfar S, Fatima I, Aslam M, Musharraf SG, Sherman NE, Moskaluk C, et al. Identification of actin beta-like 2 (ACTBL2) as novel, upregulated protein in colorectal cancer. J Proteomics. 2017;152:33-40.
64. Hao J-J, Zhi X, Wang Y, Zhang Z, Hao Z, Ye R, et al. Comprehensive Proteomic Characterization of the Human Colorectal Carcinoma Reveals Signature Proteins and Perturbed Pathways. Scientific reports [Internet]. 2017 2017/02//; 7:[42436 p.]. Available from: http://europepmc.org/abstract/MED/28181595
https://doi.org/10.1038/srep42436
https://europepmc.org/articles/PMC5299448
https://europepmc.org/articles/PMC5299448?pdf=render.
65. Quesada-Calvo F, Massot C, Bertrand V, Longuespée R, Blétard N, Somja J, et al. OLFM4, KNG1 and Sec24C identified by proteomics and immunohistochemistry as potential markers of early colorectal cancer stages. Clin Proteomics. 2017;14:9.
66. Yamamoto T, Kudo M, Peng WX, Takata H, Takakura H, Teduka K, et al. Identification of aldolase A as a potential diagnostic biomarker for colorectal cancer based on proteomic analysis using formalin-fixed paraffin-embedded tissue. Tumour Biol. 2016;37(10):13595-606.
67. Ivancic MM, Megna BW, Sverchkov Y, Craven M, Reichelderfer M, Pickhardt PJ, et al. Noninvasive Detection of Colorectal Carcinomas Using Serum Protein Biomarkers. J Surg Res. 2020;246:160-9.
68. Bhardwaj M, Gies A, Weigl K, Tikk K, Benner A, Schrotz-King P, et al. Evaluation and Validation of Plasma Proteins Using Two Different Protein Detection Methods for Early Detection of Colorectal Cancer. Cancers (Basel). 2019;11(10).
69. Yu J, Zhai X, Li X, Zhong C, Guo C, Yang F, et al. Identification of MST1 as a potential early detection biomarker for colorectal cancer through a proteomic approach. Scientific reports. 2017;7(1):14265.
70. Fan NJ, Chen HM, Song W, Zhang ZY, Zhang MD, Feng LY, et al. Macrophage mannose receptor 1 and S100A9 were identified as serum diagnostic biomarkers for colorectal cancer through a label-free quantitative proteomic analysis. Cancer Biomark. 2016;16(2):235-43.
71. Wang D-R. Hypermethylated SFRP2 gene in fecal DNA is a high potential biomarker for colorectal cancer noninvasive screening. World Journal of Gastroenterology. 2008;14(4):524.
72. Huang Z, Huang D, Ni S, Peng Z, Sheng W, Du X. Plasma microRNAs are promising novel biomarkers for early detection of colorectal cancer. International journal of cancer. 2010;127(1):118-26.

73. Wang Q, Huang Z, Ni S, Xiao X, Xu Q, Wang L, et al. Plasma miR-601 and miR-760 are novel biomarkers for the early detection of colorectal cancer. PLoS One. 2012;7(9):e44398.
74. Lee BB, Lee EJ, Jung EH, Chun H-K, Chang DK, Song SY, et al. Aberrant Methylation of APC, MGMT, RASSF2A, and Wif-1 Genes in Plasma as a Biomarker for Early Detection of Colorectal Cancer. Clinical Cancer Research. 2009;15(19):6185-91.
75. Molgaard B, Viitanen AK, Kangas A, Huhtiniemi M, Larsen ST, Vanhala E, et al. Exposure to airborne particles and volatile organic compounds from polyurethane molding, spray painting, lacquering, and gluing in a workshop. International journal of environmental research and public health. 2015;12(4):3756-73.
76. Smith PA, Sheely MV, Hakspiel SJ, Miller S. Volatile organic compounds produced during irradiation of mail. AIHA journal : a journal for the science of occupational and environmental health and safety. 2003;64(2):189-95.
77. Xiao Z, Liu S, Gu Y, Xu N, Shang Y, Zhu J. Discrimination of cherry wines based on their sensory properties and aromatic fingerprinting using HS-SPME-GC-MS and multivariate analysis. Journal of food science. 2014;79(3):C284-94.
78. Pauling L, Robinson AB, Teranishi R, Cary P. Quantitative analysis of urine vapor and breath by gas-liquid partition chromatography. Proceedings of the National Academy of Sciences of the United States of America. 1971;68(10):2374-6.
79. Phillips M, Herrera J, Krishnan S, Zain M, Greenberg J, Cataneo RN. Variation in volatile organic compounds in the breath of normal humans. J Chromatogr B Biomed Sci Appl. 1999;729(1-2):75-88.
80. Manolis A. The diagnostic potential of breath analysis. Clin Chem. 1983;29(1):5-15.
81. Sonoda H, Kohnoe S, Yamazato T, Satoh Y, Morizono G, Shikata K, et al. Colorectal cancer screening with odour material by canine scent detection. Gut. 2011;60(6):814-9.
82. Filipiak W, Mochalski P, Filipiak A, Ager C, Cumeras R, Davis CE, et al. A Compendium of Volatile Organic Compounds (VOCs) Released By Human Cell Lines. Curr Med Chem. 2016;23(20):2112-31.
83. Filipiak W, Filipiak A, Sponring A, Schmid T, Zelger B, Ager C, et al. Comparative analyses of volatile organic compounds (VOCs) from patients, tumors and transformed cell lines for the validation of lung cancer-derived breath markers. J Breath Res. 2014;8(2):027111.
84. Altomare DF, Di Lena M, Porcelli F, Trizio L, Travaglio E, Tutino M, et al. Exhaled volatile organic compounds identify patients with colorectal cancer. The British journal of surgery. 2013;100(1):144-50.
85. Haick H, Broza YY, Mochalski P, Ruzsanyi V, Amann A. Assessment, origin, and implementation of breath volatile cancer markers. Chem Soc Rev. 2014;43(5):1423-49.
86. Adam ME, Fehervari M, Boshier PR, Chin S-T, Lin G-P, Romano A, et al. Mass-Spectrometry Analysis of Mixed-Breath, Isolated-Bronchial-Breath, and Gastric-Endoluminal-Air Volatile Fatty Acids in Esophagogastric Cancer. Analytical Chemistry. 2019;91(5):3740-6.
87. Hakim M, Broza YY, Barash O, Peled N, Phillips M, Amann A, et al. Volatile organic compounds of lung cancer and possible biochemical pathways. Chem Rev. 2012;112(11):5949-66.
88. Kneepkens CM, Lepage G, Roy CC. The potential of the hydrocarbon breath test as a measure of lipid peroxidation. Free radical biology & medicine. 1994;17(2):127-60.
89. Toyokuni S. Molecular mechanisms of oxidative stress-induced carcinogenesis: From epidemiology to oxygenomics. IUBMB Life. 2008;60(7):441-7.

90. GA Burdock GF. Fenaroli's Handbook of Flavour Ingredients. 6th ed. Boca Raton: CRC Press/Taylor & Francis group; 2010.
91. Musa-Veloso K, Likhodii SS, Rarama E, Benoit S, Liu YM, Chartrand D, et al. Breath acetone predicts plasma ketone bodies in children with epilepsy on a ketogenic diet. Nutrition. 2006;22(1):1-8.
92. Spaněl P, Dryahina K, Rejšková A, Chippendale TW, Smith D. Breath acetone concentration; biological variability and the influence of diet. Physiol Meas. 2011;32(8):N23-31.
93. Weinburg R. The biology of cancer. New York: Garland publishing inc.; 2006.
94. Gatenby RA, Gillies RJ. Why do cancers have high aerobic glycolysis? Nat Rev Cancer. 2004;4(11):891-9.
95. de Lacy Costello B, Amann A, Al-Kateb H, Flynn C, Filipiak W, Khalid T, et al. A review of the volatiles from the healthy human body. J Breath Res. 2014;8(1):014001.
96. van Keulen KE, Jansen ME, Schrauwen RWM, Kolkman JJ, Siersema PD. Volatile organic compounds in breath can serve as a non-invasive diagnostic biomarker for the detection of advanced adenomas and colorectal cancer. Aliment Pharmacol Ther. 2020;51(3):334-46.
97. Atreja A, Fu AZ, Sanaka MR, Vargo JJ. Non-invasive testing for Helicobacter pylori in patients hospitalized with peptic ulcer hemorrhage: a cost-effectiveness analysis. Digestive diseases and sciences. 2010;55(5):1356-63.
98. Gahleitner F, Guallar-Hoyas C, Beardsmore CS, Pandya HC, Thomas CP. Metabolomics pilot study to identify volatile organic compound markers of childhood asthma in exhaled breath. Bioanalysis. 2013;5(18):2239-47.
99. Harding P, Field PH. Breathalyzer accuracy in actual law enforcement practice: a comparison of blood- and breath-alcohol results in Wisconsin drivers. Journal of forensic sciences. 1987;32(5):1235-40.
100. Petrone P, Sarkisyan G, Fernandez M, Coloma E, Akopian G, Ortega A, et al. Small intestinal bacterial overgrowth in patients with lower gastrointestinal symptoms and a history of previous abdominal surgery. Archives of surgery (Chicago, Ill : 1960). 2011;146(4):444-7.
101. Robroeks CM, van Berkel JJ, Jobsis Q, van Schooten FJ, Dallinga JW, Wouters EF, et al. Exhaled volatile organic compounds predict exacerbations of childhood asthma in a 1-year prospective study. The European respiratory journal. 2013;42(1):98-106.
102. Hanna GB, Boshier PR, Markar SR, Romano A. Accuracy and Methodologic Challenges of Volatile Organic Compound-Based Exhaled Breath Tests for Cancer Diagnosis: A Systematic Review and Meta-analysis. JAMA oncology. 2019;5:e182815.
103. Kumar S, Huang J, Abbassi-Ghadi N, Spanel P, Smith D, Hanna GB. Selected ion flow tube mass spectrometry analysis of exhaled breath for volatile organic compound profiling of esophago-gastric cancer. Analytical chemistry. 2013;85(12):6121-8.
104. Kumar S, Huang J, Abbassi-Ghadi N, Mackenzie HA, Veselkov KA, Hoare JM, et al. Mass Spectrometric Analysis of Exhaled Breath for the Identification of Volatile Organic Compound Biomarkers in Esophageal and Gastric Adenocarcinoma. Annals of surgery. 2015;262(6):981-90.
105. Markar SR, Lagergren J, Hanna GB. Research protocol for a diagnostic study of non-invasive exhaled breath analysis for the prediction of oesophago-gastric cancer. BMJ Open. 2016;6(1).

106. Adam M. Role of microbiome in volatile production in oesophageal adenocarcinoma: Imperial College London; 2015-2018.
107. Antonowicz S. Aldehyde dysfunction in oesophageal adenocarcinoma-PhD: Imperial College London; 2013-2016.
108. Wiggins T. Phenol and aromatic acid dysfunction in oesophageal adenocarcinoma.: Imperial College London; 2013-2016.
109. Markar SR, Chin ST, Romano A, Wiggins T, Antonowicz S, Paraskeva P, et al. Breath Volatile Organic Compound Profiling of Colorectal Cancer Using Selected Ion Flow-tube Mass Spectrometry. Annals of Surgery. 2019;269(5):903-10.
110. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open. 2016;6(11):e012799.
111. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Annals of internal medicine. 2011;155(8):529-36.
112. Batty CA, Cauchi M, Lourenço C, Hunter JO, Turner C. Use of the Analysis of the Volatile Faecal Metabolome in Screening for Colorectal Cancer. PLoS One. 2015;10(6):e0130301.
113. Ahmed I, Greenwood R, Costello Bde L, Ratcliffe NM, Probert CS. An investigation of fecal volatile organic metabolites in irritable bowel syndrome. PLoS One. 2013;8(3):e58204.
114. Hicks LC, Huang J, Kumar S, Powles ST, Orchard TR, Hanna GB, et al. Analysis of Exhaled Breath Volatile Organic Compounds in Inflammatory Bowel Disease: A Pilot Study. Journal of Crohn's and Colitis. 2015;9(9):731-7.
115. Monasta L, Pierobon C, Princivalle A, Martelossi S, Marcuzzi A, Pasini F, et al. Inflammatory bowel disease and patterns of volatile organic compounds in the exhaled breath of children: A case-control study using Ion Molecule Reaction-Mass Spectrometry. PLoS One. 2017;12(8):e0184118.
116. Altomare DF, Di Lena M, Porcelli F, Travaglio E, Longobardi F, Tutino M, et al. Effects of Curative Colorectal Cancer Surgery on Exhaled Volatile Organic Compounds and Potential Implications in Clinical Follow-up. Annals of surgery. 2015;262(5):862-6; discussion 6-7.
117. Amal H, Leja M, Funka K, Lasina I, Skapars R, Sivins A, et al. Breath testing as potential colorectal cancer screening tool. International Journal of Cancer. 2016;138(1):229-36.
118. Rieder F, Kurada S, Grove D, Cikach F, Lopez R, Patel N, et al. A Distinct Colon-Derived Breath Metabolome is Associated with Inflammatory Bowel Disease, but not its Complications. Clinical and Translational Gastroenterology. 2016;7(11):e201.
119. Bodelier AG, Smolinska A, Baranska A, Dallinga JW, Mujagic Z, Vanhees K, et al. Volatile Organic Compounds in Exhaled Air as Novel Marker for Disease Activity in Crohn's Disease: A Metabolomic Approach. Inflamm Bowel Dis. 2015;21(8):1776-85.
120. Patel N, Alkhouri N, Eng K, Cikach F, Mahajan L, Yan C, et al. Metabolomic analysis of breath volatile organic compounds reveals unique breathprints in children with inflammatory bowel disease: a pilot study. Alimentary Pharmacology & Therapeutics. 2014;40(5):498-507.
121. Dryahina K, Smith D, Bortlík M, Machková N, Lukáš M, Španěl P. Pentane and other volatile organic compounds, including carboxylic acids, in the exhaled breath of patients with Crohn's disease and ulcerative colitis. J Breath Res. 2017;12(1):016002.

122. Smolinska A, Bodelier AG, Dallinga JW, Masclee AA, Jonkers DM, van Schooten FJ, et al. The potential of volatile organic compounds for the detection of active disease in patients with ulcerative colitis. Aliment Pharmacol Ther. 2017;45(9):1244-54.
123. Walton C, Fowler DP, Turner C, Jia W, Whitehead RN, Griffiths L, et al. Analysis of volatile organic compounds of bacterial origin in chronic gastrointestinal diseases. Inflamm Bowel Dis. 2013;19(10):2069-78.
124. Ahmed I, Greenwood R, Costello B, Ratcliffe N, Probert CS. Investigation of faecal volatile organic metabolites as novel diagnostic biomarkers in inflammatory bowel disease. Aliment Pharmacol Ther. 2016;43(5):596-611.
125. Bond A, Greenwood R, Lewis S, Corfe B, Sarkar S, O'Toole P, et al. Volatile organic compounds emitted from faeces as a biomarker for colorectal cancer. Alimentary Pharmacology & Therapeutics. 2019;49(8):1005-12.
126. Kim S, Yin X, Prodhan MAI, Zhang X, Zhong Z, Kato I. Global Plasma Profiling for Colorectal Cancer-Associated Volatile Organic Compounds: a Proof-of-Principle Study. Journal of Chromatographic Science. 2019;57(5):385-96.
127. Wang C, Li P, Lian A, Sun B, Wang X, Guo L, et al. Blood volatile compounds as biomarkers for colorectal cancer. Cancer Biology & Therapy. 2014;15(2):200-6.
128. Wang C, Ke C, Wang X, Chi C, Guo L, Luo S, et al. Noninvasive detection of colorectal cancer by analysis of exhaled breath. Analytical & Bioanalytical Chemistry. 2014;406(19):4757-63.
129. Crotti S, Agnoletto E, Cancemi G, Di Marco V, Traldi P, Pucciarelli S, et al. Altered plasma levels of decanoic acid in colorectal cancer as a new diagnostic biomarker. Analytical and Bioanalytical Chemistry. 2016;408(23):6321-8.
130. Tan B, Qiu Y, Zou X, Chen T, Xie G, Cheng Y, et al. Metabonomics identifies serum metabolite markers of colorectal cancer. Journal of Proteome Research. 2013;12(6):3000-9.
131. Nishiumi S, Kobayashi T, Ikeda A, Yoshie T, Kibi M, Izumi Y, et al. A novel serum metabolomics-based diagnostic approach for colorectal cancer. PLoS One. 2012;7(7):e40459.
132. Cheng Y, Xie G, Chen T, Qiu Y, Zou X, Zheng M, et al. Distinct urinary metabolic profile of human colorectal cancer. Journal of Proteome Research. 2012;11(2):1354-63.
133. Baranska A, Tigchelaar E, Smolinska A, Dallinga JW, Moonen EJC, Dekens JAM, et al. Profile of volatile organic compounds in exhaled breath changes as a result of gluten-free diet. Journal of breath research. 2013;7(3):037104.
134. King J, Mochalski P, Kupferthaler A, Unterkofler K, Koc H, Filipiak W, et al. Dynamic profiles of volatile organic compounds in exhaled breath as determined by a coupled PTR-MS/GC-MS study. Physiol Meas. 2010;31(9):1169-84.
135. Widlak MM, Neal M, Daulton E, Thomas CL, Tomkins C, Singh B, et al. Risk stratification of symptomatic patients suspected of colorectal cancer using faecal and urinary markers. Colorectal Disease. 2018;20(12):O335-O42.
136. Spanel P, Smith D. Selected ion flow tube mass spectrometry for on-line trace gas analysis in biology and medicine. European journal of mass spectrometry (Chichester, England). 2007;13(1):77-82.
137. Gilchrist FJ, Razavi C, Webb AK, Jones AM, Spaněl P, Smith D, et al. An investigation of suitable bag materials for the collection and storage of breath samples containing hydrogen cyanide. J Breath Res. 2012;6(3):036004.
138. Harshman SW, Mani N, Geier BA, Kwak J, Shepard P, Fan M, et al. Storage stability of exhaled breath on Tenax TA. J Breath Res. 2016;10(4):046008.

139. Kang S, Paul Thomas CL. How long may a breath sample be stored for at -80 degrees C? A study of the stability of volatile organic compounds trapped onto a mixed Tenax:Carbograph trap adsorbent bed from exhaled breath. J Breath Res. 2016;10(2):026011.
140. Woolfenden E. Monitoring VOCs in Air Using Sorbent Tubes Followed by Thermal Desorption-Capillary GC Analysis: Summary of Data and Practical Guidelines. Journal of the Air & Waste Management Association. 1997;47(1):20-36.
141. Markes International. Thermal Desorption Technical Support document, Sorbent tube storage. 2018 [Available from: https://www.chromservis.eu/i/sorbent-tube-storage/c/thermal-desorption.
142. Romano A, Doran S, Belluomo I, Hanna GB. High-Throughput Breath Volatile Organic Compound Analysis Using Thermal Desorption Proton Transfer Reaction Time-of-Flight Mass Spectrometry. Analytical chemistry. 2018;90(17):10204-10.
143. Doran SLF, Romano A, Hanna GB. Optimisation of sampling parameters for standardised exhaled breath sampling. J Breath Res. 2017;12(1):016007.
144. Harshman SW, Pitsch RL, Davidson CN, Scott AM, Hill EM, Smith ZK, et al. Characterization of standardized breath sampling for off-line field use. J Breath Res. 2019;14(1):016009.
145. Phillips M, Gleeson K, Hughes JM, Greenberg J, Cataneo RN, Baker L, et al. Volatile organic compounds in breath as markers of lung cancer: a cross-sectional study. Lancet. 1999;353(9168):1930-3.
146. Filipiak W, Sponring A, Filipiak A, Ager C, Schubert J, Miekisch W, et al. TD-GC-MS analysis of volatile metabolites of human lung cancer and normal cells in vitro. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2010;19(1):182-95.
147. Phillips M, Cataneo RN, Ditkoff BA, Fisher P, Greenberg J, Gunawardena R, et al. Prediction of breast cancer using volatile biomarkers in the breath. Breast cancer research and treatment. 2006;99(1):19-21.
148. Dadamio J, Van den Velde S, Laleman W, Van Hee P, Coucke W, Nevens F, et al. Breath biomarkers of liver cirrhosis. Journal of chromatography B, Analytical technologies in the biomedical and life sciences. 2012;905:17-22.
149. Khalid TY, Costello BDL, Ewen R, White P, Stevens S, Gordon F, et al. Breath volatile analysis from patients diagnosed with harmful drinking, cirrhosis and hepatic encephalopathy: a pilot study. Metabolomics. 2013;9(5):938-48.
150. Chin S-T, Romano A, Doran SLF, Hanna GB. Cross-platform mass spectrometry annotation in breathomics of oesophageal-gastric cancer. Scientific reports. 2018;8(1):5139.
151. Spanel P, Smith D. Progress in SIFT-MS: breath analysis and other applications. Mass spectrometry reviews. 2011;30(2):236-67.
152. Spanel P, Smith D, Holland TA, Al Singary W, Elder JB. Analysis of formaldehyde in the headspace of urine from bladder and prostate cancer patients using selected ion flow tube mass spectrometry. Rapid communications in mass spectrometry : RCM. 1999;13(14):1354-9.
153. Hashoul D, Haick H. Sensors for detecting pulmonary diseases from exhaled breath. European Respiratory Review. 2019;28(152):190011.
154. Vishinkin R, Haick H. Nanoscale Sensor Technologies for Disease Detection via Volatolomics. Small. 2015;11(46):6142-64.

155. Montuschi P, Mores N, Trové A, Mondino C, Barnes PJ. The electronic nose in respiratory medicine. Respiration. 2013;85(1):72-84.
156. Tang Z, Yang J, Yu J, Cui B. A colorimetric sensor for qualitative discrimination and quantitative detection of volatile amines. Sensors (Basel). 2010;10(7):6463-76.
157. Amal H, Shi DY, Ionescu R, Zhang W, Hua QL, Pan YY, et al. Assessment of ovarian cancer conditions from exhaled breath. International journal of cancer. 2015;136(6):E614-22.
158. Peng G, Hakim M, Broza YY, Billan S, Abdah-Bortnyak R, Kuten A, et al. Detection of lung, breast, colorectal, and prostate cancers from exhaled breath using a single array of nanosensors. British Journal of Cancer. 2010;103(4):542-51.
159. Herbig J, Beauchamp J. Towards standardization in the analysis of breath gas volatiles. J Breath Res. 2014;8(3):037101.
160. Mochalski P, Sponring A, King J, Unterkofler K, Troppmair J, Amann A. Release and uptake of volatile organic compounds by human hepatocellular carcinoma cells (HepG2) in vitro. Cancer Cell Int. 2013;13(1):72.
161. Diskin AM, Spanel P, Smith D. Increase of acetone and ammonia in urine headspace and breath during ovulation quantified using selected ion flow tube mass spectrometry. Physiol Meas. 2003;24(1):191-9.
162. Erhart S, Amann A, Haberlandt E, Edlinger G, Schmid A, Filipiak W, et al. 3-Heptanone as a potential new marker for valproic acid therapy. J Breath Res. 2009;3(1):016004.
163. Markar SR, Wiggins T, Antonowicz S, Chin ST, Romano A, Nikolic K, et al. Assessment of a Noninvasive Exhaled Breath Test for the Diagnosis of Oesophagogastric Cancer. JAMA oncology. 2018;4(7):970-6.
164. Pysanenko A, Spaněl P, Smith D. A study of sulfur-containing compounds in mouth- and nose-exhaled breath and in the oral cavity using selected ion flow tube mass spectrometry. J Breath Res. 2008;2(4):046004.
165. Shen VK, Siderius, D.W., Krekelberg, W.P., and Hatch, H.W. NIST Standard Reference Simulation Website. Gaithersburg MD: National Institute of Standards and Technology, ; [
166. R Core Team: R: a language and environment for statistical computing, 2017 [Available from: https://www.R-project.org/.
167. Romano A, Hanna GB. Identification and quantification of VOCs by proton transfer reaction time of flight mass spectrometry: An experimental workflow for the optimization of specificity, sensitivity, and accuracy. J Mass Spectrom. 2018;53(4):287-95.
168. Smith D, Spanel P. Pitfalls in the analysis of volatile breath biomarkers: suggested solutions and SIFT-MS quantification of single metabolites. J Breath Res. 2015;9(2):022001.
169. Chin ST, Romano A, Doran SLF, Hanna GB. Cross-platform mass spectrometry annotation in breathomics of oesophageal-gastric cancer. Scientific reports. 2018;8(1):5139.
170. Turner C, Spanel P, Smith D. A longitudinal study of ethanol and acetaldehyde in the exhaled breath of healthy volunteers using selected-ion flow-tube mass spectrometry. Rapid Commun Mass Spectrom. 2006;20(1):61-8.
171. Aksenov AA, Laponogov I, Zhang Z, Doran SLF, Belluomo I, Veselkov D, et al. Algorithmic Learning for Auto-deconvolution of GC-MS Data to Enable Molecular Networking within GNPS. bioRxiv. 2020:2020.01.13.905091.
172. Aksenov AA, Laponogov I, Zhang Z, Doran SLF, Belluomo I, Veselkov D, et al. Auto-deconvolution and molecular networking of gas chromatography–mass spectrometry data. Nature Biotechnology. 2021;39:169-73.

173. Lewis C, Rieman J. Task-centered User Interface Design: A Practical Introduction: University of Colorado, Boulder, Department of Computer Science; 1993.
174. Atkin WS, Saunders BP. Surveillance guidelines after removal of colorectal adenomatous polyps. Gut. 2002;51(suppl 5):v6-v9.
175. Corp I. IBM SPSS Statistics for Mac. 25.0 ed. Armonk, NY: IBM Corp.; 2017.
176. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011;12(null):2825–30.
177. National Center for Biotechnology Information. Pubchem Compound Summary for CID 8252, Propyl-propionate: National Centre for Biotechnology Information; 2020 [Available from: https://pubchem.ncbi.nlm.nih.gov/compound/Propyl-propionate.
178. Mochalski P, Leja M, Gasenko E, Skapars R, Santare D, Sivins A, et al. Ex vivo emission of volatile organic compounds from gastric cancer and non-cancerous tissue. Journal of breath research. 2018;12(4):046005.
179. Bouza M, Gonzalez-Soto J, Pereiro R, de Vicente JC, Sanz-Medel A. Exhaled breath and oral cavity VOCs as potential biomarkers in oral cancer patients. J Breath Res. 2017;11(1):016015.
180. Tong H, Wang Y, Li Y, Liu S, Chi C, Liu D, et al. Volatile organic metabolites identify patients with gastric carcinoma, gastric ulcer, or gastritis and control patients. Cancer Cell International. 2017;17(1):108.
181. O'Neil MJ HP, Koch CB, and Roman KJ. The Merck Index: An Encyclopedia of Chemicals, Drugs, and Biologicals, 14th ed. Edited by Journal of the American Chemical Society. 2007;127(7):2197.
182. Gaida A, Holz O, Nell C, Schuchardt S, Lavae-Mokhtari B, Kruse L, et al. A dual center study to compare breath volatile organic compounds from smokers and non-smokers with and without COPD. Journal of breath research. 2016;10(2):026006.
183. Phillips M, Cataneo RN, Saunders C, Hope P, Schmitt P, Wai J. Volatile biomarkers in the breath of women with breast cancer. J Breath Res. 2010;4(2):026003.
184. Giavalisco P, Hummel J, Lisec J, Inostroza AC, Catchpole G, Willmitzer L. High-Resolution Direct Infusion-Based Mass Spectrometry in Combination with Whole 13C Metabolome Isotope Labeling Allows Unambiguous Assignment of Chemical Sum Formulas. Analytical Chemistry. 2008;80(24):9417-25.
185. Lemar KM, Passa O, Aon MA, Cortassa S, Müller CT, Plummer S, et al. Allyl alcohol and garlic (Allium sativum) extract produce oxidative stress in Candida albicans. Microbiology (Reading). 2005;151(Pt 10):3257-65.
186. Leffingwell JC AE. Volatile constituents of Perique tobacco. Journal of Environmental, Agricultural and Food Chemistry. 2005;4(2):899-915.
187. National Center for Biotechnology Information. PubChem Compound Summary for 2_3_4-Trimethyl-3-pentanol: National Center for Biotechnology Information; 2020 [Available from: https://pubchem.ncbi.nlm.nih.gov/compound/Ethyl-methylphosphinate.
188. Wishart DS, Tzur D, Knox C, Eisner R, Guo AC, Young N, et al. HMDB: the Human Metabolome Database. Nucleic Acids Res. 2007;35(Database issue):D521-6.
189. National Center for Biotechnology Information. PubChem Compound Summary for CID 6328184, Ethyl methylphosphinate: National Center for Biotechnology Information; 2020 [Available from: https://pubchem.ncbi.nlm.nih.gov/compound/Ethyl-methylphosphinate.

190. Kaur G, Verma N. Nature curing cancer - review on structural modification studies with natural active compounds having anti-tumor efficiency. Biotechnol Rep (Amst). 2015;6:64-78.
191. Packard H. "Safety Data Sheet" (PDF) UK: Hewlett Packard; 2014 [Available from: https://en.wikipedia-on-ipfs.org/wiki/2-pyrrolidone.html.
192. Peng G, Tisch U, Adams O, Hakim M, Shehada N, Broza YY, et al. Diagnosing lung cancer in exhaled breath using gold nanoparticles. Nature Nanotechnology. 2009;4(10):669-73.
193. Performance Characteristics of Fecal Immunochemical Tests for Colorectal Cancer and Advanced Adenomatous Polyps. Annals of internal medicine. 2019;170(5):319-29.
194. European Molecular Biology Laboratory European Bioinformatics Institute. CHEBI:89828 - propyl propionate 2017 [Available from: https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:89828.
195. Garner CE, Smith S, de Lacy Costello B, White P, Spencer R, Probert CS, et al. Volatile organic compounds from feces and their potential for diagnosis of gastrointestinal disease. Faseb j. 2007;21(8):1675-88.
196. Raman M, Ahmed I, Gillevet PM, Probert CS, Ratcliffe NM, Smith S, et al. Fecal microbiome and volatile organic compound metabolome in obese humans with nonalcoholic fatty liver disease. Clin Gastroenterol Hepatol. 2013;11(7):868-75.e1-3.
197. De Angelis M, Vannini L, Di Cagno R, Cavallo N, Minervini F, Francavilla R, et al. Salivary and fecal microbiota and metabolome of celiac children under gluten-free diet. Int J Food Microbiol. 2016;239:125-32.
198. Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. Host-bacterial mutualism in the human intestine. Science. 2005;307(5717):1915-20.
199. Vinolo MA, Rodrigues HG, Nachbar RT, Curi R. Regulation of inflammation by short chain fatty acids. Nutrients. 2011;3(10):858-76.
200. Takaishi H, Matsuki T, Nakazawa A, Takada T, Kado S, Asahara T, et al. Imbalance in intestinal microflora constitution could be involved in the pathogenesis of inflammatory bowel disease. Int J Med Microbiol. 2008;298(5-6):463-72.
201. Huda-Faujan N, Abdulamir AS, Fatimah AB, Anas OM, Shuhaimi M, Yazid AM, et al. The impact of the level of the intestinal short chain Fatty acids in inflammatory bowel disease patients versus healthy subjects. Open Biochem J. 2010;4:53-8.
202. Mortensen PB, Clausen MR. Short-chain fatty acids in the human colon: relation to gastrointestinal health and disease. Scand J Gastroenterol Suppl. 1996;216:132-48.
203. Imai T. Human carboxylesterase isozymes: catalytic properties and rational drug design. Drug Metab Pharmacokinet. 2006;21(3):173-85.
204. Zhu HJ, Wang X, Gawronski BE, Brinda BJ, Angiolillo DJ, Markowitz JS. Carboxylesterase 1 as a determinant of clopidogrel metabolism and activation. J Pharmacol Exp Ther. 2013;344(3):665-72.
205. Wang D, Zou L, Jin Q, Hou J, Ge G, Yang L. Human carboxylesterases: a comprehensive review. Acta Pharm Sin B. 2018;8(5):699-712.
206. Buszewski B, Ulanowska A, Ligor T, Jackowski M, Kłodzińska E, Szeliga J. Identification of volatile organic compounds secreted from cancer tissues and bacterial cultures. Journal of chromatography B, Analytical technologies in the biomedical and life sciences. 2008;868(1-2):88-94.
207. Bos LD, Sterk PJ, Schultz MJ. Volatile metabolites of pathogens: a systematic review. PLoS Pathog. 2013;9(5):e1003311.

208. Tangerman A. Measurement and biological significance of the volatile sulfur compounds hydrogen sulfide, methanethiol and dimethyl sulfide in various biological matrices. Journal of chromatography B, Analytical technologies in the biomedical and life sciences. 2009;877(28):3366-77.
209. Pimenta MJ, Kaneta T, Larondelle Y, Dohmae N, Kamiya Y. S-adenosyl-L-methionine:L-methionine S-methyltransferase from germinating barley. Purification and localization. Plant Physiol. 1998;118(2):431-8.
210. Tangerman A, Meuwese-Arends MT, Jansen JB. Cause and composition of foetor hepaticus. Lancet. 1994;343(8895):483.
211. Suarez FL, Springfield J, Levitt MD. Identification of gases responsible for the odour of human flatus and evaluation of a device purported to reduce this odour. Gut. 1998;43(1):100-4.
212. Pitcher MCL, Beatty ER, Cummings JH. The contribution of sulphate reducing bacteria and 5-aminosalicylic acid to faecal sulphide in patients with ulcerative colitis. Gut. 2000;46(1):64.
213. Vogelstein BaK, K. W. The Genetic Basis of Human Cancer. New York.: McGraw-Hill,; 1998.
214. Preedy VR. Index. In: Preedy VR, editor. Aging. San Diego: Academic Press; 2014. p. 297-301.
215. Rice-Evans CA. Chapter 5 - Formation of free radicals and mechanisms of action in normal biochemical processes and pathological states. In: Rice-Evans CA, Burdon RH, editors. New Comprehensive Biochemistry. 28: Elsevier; 1994. p. 131-53.
216. O'Neill HJ, Gordon SM, O'Neill MH, Gibbons RD, Szidon JP. A computerized classification technique for screening for the presence of breath biomarkers in lung cancer. Clin Chem. 1988;34(8):1613-8.
217. Risby TH, Sehnert SS. Clinical application of breath biomarkers of oxidative stress status. Free radical biology & medicine. 1999;27(11-12):1182-92.
218. Singh S.N. KB, Mishra S. Microbial Degradation of Alkanes. Berlin, Heidelberg: Springer; 2012.
219. Burdock GA. Encyclopedia of Food and Color Additives: CRC Press; 1997.
220. Seitz HK, Egerer G, Simanowski UA, Waldherr R, Eckey R, Agarwal DP, et al. Human gastric alcohol dehydrogenase activity: effect of age, sex, and alcoholism. Gut. 1993;34(10):1433-7.
221. Sofer W, Martin PF. Analysis of alcohol dehydrogenase gene expression in Drosophila. Annu Rev Genet. 1987;21:203-25.
222. Smolinska A, Tedjo DI, Blanchet L, Bodelier A, Pierik MJ, Masclee AAM, et al. Volatile metabolites in breath strongly correlate with gut microbiome in CD patients. Analytica Chimica Acta. 2018;1025:1-11.
223. Kumar S, Huang J, Cushnir JR, Spanel P, Smith D, Hanna GB. Selected ion flow tube-MS analysis of headspace vapor from gastric content for the diagnosis of gastro-esophageal cancer. Analytical chemistry. 2012;84(21):9550-7.
224. Huang J, Kumar S, Abbassi-Ghadi N, Spanel P, Smith D, Hanna GB. Selected ion flow tube mass spectrometry analysis of volatile metabolites in urine headspace for the profiling of gastro-esophageal cancer. Analytical chemistry. 2013;85(6):3409-16.
225. Filipiak W, Ruzsanyi V, Mochalski P, Filipiak A, Bajtarevic A, Ager C, et al. Dependence of exhaled breath composition on exogenous factors, smoking habits and exposure to air pollutants. J Breath Res. 2012;6(3):036008.

226. Lee JK, Liles EG, Bent S, Levin TR, Corley DA. Accuracy of fecal immunochemical tests for colorectal cancer: systematic review and meta-analysis. Annals of internal medicine. 2014;160(3):171.
227. Mowat C, Digby J, Strachan JA, Wilson R, Carey FA, Fraser CG, et al. Faecal haemoglobin and faecal calprotectin as indicators of bowel disease in patients presenting to primary care with bowel symptoms. Gut. 2016;65(9):1463.
228. Rodríguez-Alonso L, Rodríguez-Moranta F, Ruiz-Cerulla A, Lobatón T, Arajol C, Binefa G, et al. An urgent referral strategy for symptomatic patients with suspected colorectal cancer based on a quantitative immunochemical faecal occult blood test. Dig Liver Dis. 2015;47(9):797-804.
229. McDonald PJ, Digby J, Innes C, Strachan JA, Carey FA, Steele RJ, et al. Low faecal haemoglobin concentration potentially rules out significant colorectal disease. Colorectal Dis. 2013;15(3):e151-9.
230. Terhaar sive Droste JS, Oort FA, van der Hulst RW, van Heukelem HA, Loffeld RJ, van Turenhout ST, et al. Higher fecal immunochemical test cutoff levels: lower positivity rates but still acceptable detection rates for early-stage colorectal cancers. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2011;20(2):272-80.
231. Godber IM, Todd LM, Fraser CG, MacDonald LR, Younes HB. Use of a faecal immunochemical test for haemoglobin can aid in the investigation of patients with lower abdominal symptoms. Clin Chem Lab Med. 2016;54(4):595-602.
232. Cross AJ, Wooldrage K, Robbins EC, Kralj-Hans I, MacRae E, Piggott C, et al. Faecal immunochemical tests (FIT) versus colonoscopy for surveillance after screening and polypectomy: a diagnostic accuracy and cost-effectiveness study. Gut. 2018:gutjnl-2018-317297.
233. Lecky DM, Hawking MKD, McNulty CAM, group Es. Patients' perspectives on providing a stool sample to their GP: a qualitative study. Br J Gen Pract. 2014;64(628):e684-e93.
234. Hirst Y, Stoffel S, Baio G, McGregor L, von Wagner C. Uptake of the English Bowel (Colorectal) Cancer Screening Programme: an update 5 years after the full roll-out. Eur J Cancer. 2018;103:267-73.
235. Cancer Intelligence Statistical Information Team at Cancer Research UK. Bowel cancer; risks and causes UK2018 [Available from: https://www.cancerresearchuk.org/about-cancer/bowel-cancer/risks-causes.
236. Boshier PR, Fehervari M, Markar SR, Purkayastha S, Spanel P, Smith D, et al. Variation in Exhaled Acetone and Other Ketones in Patients Undergoing Bariatric Surgery: a Prospective Cross-sectional Study. Obesity Surgery. 2018;28(8):2439-46.
237. Leja M, Amal H, Lasina I, Skapars R, Sivins A, Ancans G, et al. Analysis of the effects of microbiome-related confounding factors on the reproducibility of the volatolomic test. Journal of breath research. 2016;10(3):037101.
238. Smith D, Spanel P, Davies S. Trace gases in breath of healthy volunteers when fasting and after a protein-calorie meal: a preliminary study. Journal of Applied Physiology. 1999;87(5):1584-8.
239. van Oort PM, Povoa P, Schnabel R, Dark P, Artigas A, Bergmans D, et al. The potential role of exhaled breath analysis in the diagnostic process of pneumonia-a systematic review. J Breath Res. 2018;12(2):024001.

240. Capuano R, Chimenti M, Dini F, Domakoski A, Martinelli E, Conigliaro P, et al. FRI0012 Role of volatile compounds released by synovial fluid in the diagnosis of osteoarthritis and rheumatoid arthritis of the knee joint. Annals of the Rheumatic Diseases. 2017;76(Suppl 2):484-.
241. Brinkman P, Ahmed WM, Gómez C, Knobel HH, Weda H, Vink TJ, et al. Exhaled volatile organic compounds as markers for medication use in asthma. The European respiratory journal. 2020;55(2).
242. Schubert JK, Miekisch W, Birken T, Geiger K, Nöldge-Schomburg GFE. Impact of inspired substance concentrations on the results of breath analysis in mechanically ventilated patients. Biomarkers. 2005;10(2-3):138-52.
243. Boshier PR, Cushnir JR, Priest OH, Marczin N, Hanna GB. Variation in the levels of volatile trace gases within three hospital environments: implications for clinical breath testing. J Breath Res. 2010;4(3):031001.
244. Cope KA, Watson MT, Foster WM, Sehnert SS, Risby TH. Effects of ventilation on the collection of exhaled breath in humans. J Appl Physiol (1985). 2004;96(4):1371-9.
245. Wang T, Pysanenko A, Dryahina K, Spaněl P, Smith D. Analysis of breath, exhaled via the mouth and nose, and the air in the oral cavity. J Breath Res. 2008;2(3):037013.
246. Phenomenex I. GC Troubleshooting guide. 2014 [Available from: https://phenomenex.blob.core.windows.net/documents/b043f709-428f-4697-9791-7b0f2baf9b77.pdf.
247. Jeong YH, Kim KO, Park CS, Kim SB, Lee SH, Jang BI. Risk Factors of Advanced Adenoma in Small and Diminutive Colorectal Polyp. J Korean Med Sci. 2016;31(9):1426-30.
248. Brenner H, Hoffmeister M, Stegmaier C, Brenner G, Altenhofen L, Haug U. Risk of progression of advanced adenomas to colorectal cancer by age and sex: estimates based on 840,149 screening colonoscopies. Gut. 2007;56(11):1585-9.
249. Hamilton W, Green T, Martins T, Elliott K, Rubin G, Macleod U. Evaluation of risk assessment tools for suspected cancer in general practice: a cohort study. Br J Gen Pract. 2013;63(606):e30-6.
250. 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.
251. Flashman K, O'Leary DP, Senapati A, Thompson MR. The Department of Health's "two week standard" for bowel cancer: is it working? Gut. 2004;53.
252. Jones R, Rubin G, Hungin P. Is the two week rule for cancer referrals working? BMJ (Clinical research ed). 2001;322(7302):1555-6.
253. Powell J, McConkey CC. The rising trend in oesophageal adenocarcinoma and gastric cardia. European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation (ECP). 1992;1(3):265-9.
254. Blot WJ, Devesa SS, Kneller RW, Fraumeni JF, Jr. Rising incidence of adenocarcinoma of the esophagus and gastric cardia. Jama. 1991;265(10):1287-9.
255. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2012. CA: A Cancer Journal for Clinicians. 2012;62(1):10-29.
256. Kohn GP, Galanko JA, Meyers MO, Feins RH, Farrell TM. National trends in esophageal surgery--are outcomes as good as we believe? Journal of gastrointestinal surgery : official journal of the Society for Surgery of the Alimentary Tract. 2009;13(11):1900-10; discussion 10-2.

257. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. CA: A Cancer Journal for Clinicians. 2013;63(1):11-30.
258. Lorenz D, Origer J, Pauthner M, Graupe F, Fisseler-Eckhoff A, Stolte M, et al. Prognostic Risk Factors of Early Esophageal Adenocarcinomas. Annals of Surgery. 2014;259(3):469-76.
259. Lao-Sirieix P, Fitzgerald RC. Screening for oesophageal cancer. Nature reviews Clinical oncology. 2012;9(5):278-87.
260. Cancer research UK. Pancreatic cancer survival statistics 2014 [Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/pancreatic-cancer/survival#heading-Zero.
261. UK NCRASaCR. "Chemotherapy, Radiotherapy and Tumour Resections in England: 2013-2014" workbook. London: NCRAS; 2017. London: NCRAS; 2017.
262. National Institute of Clinical Excellence. Suspected cancer: recognition and referral. NICE guideline [NG12]. Upper gastrointestinal tract cancers. UK 2015 [updated July 2017] [Available from: https://www.nice.org.uk/guidance/ng12/chapter/1-Recommendations-organised-by-site-of-cancer#upper-gastrointestinal-tract-cancers.
263. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology. 2006;3(2):77-101.
264. National Institute of Health Research. Faecal immunochemical testing (FIT) for adenoma surveillance. FIT for Follow-up: National Institute of Health Research; 2019 [Available from: https://www.journalslibrary.nihr.ac.uk/programmes/hta/0922192#/.
265. "The American Heritage Dictionary entry: focus group". : Company, Houghton Mifflin Harcourt Publishing.; 2018 [Available from: www.ahdictionary.com.
266. Freeman M, Offman J, Walter FM, Sasieni P, Smith SG. Acceptability of the Cytosponge procedure for detecting Barrett's oesophagus: a qualitative study. BMJ Open. 2017;7(3):e013901.
267. Honkasalo MT, Linna M, Sane T, Honkasalo A, Elonheimo O. A comparative study of two various models of organising diabetes follow-up in public primary health care - the model influences the use of services, their quality and costs. BMC Health Serv Res. 2014;14:26.
268. Fitzsimons B, Wilton L, Lamont T, McCulloch L, Boyce J. The Audit Commission review of diabetes services in England and Wales, 1998-2001. Diabetic medicine : a journal of the British Diabetic Association. 2002;19 Suppl 4:73-8.
269. Vedsted P. Gender differences in the use of health care system. Ugeskrift for laeger. 2007;169(25):2403-8.
270. Jørgensen JT, Andersen JS, Tjønneland A, Andersen ZJ. Determinants related to gender differences in general practice utilization: Danish Diet, Cancer and Health Cohort. Scandinavian journal of primary health care. 2016;34(3):240-9.
271. Foy R, Parry J, McAvoy B. Clinical trials in primary care. Targeted payments for trials might help improve recruitment and quality. 1998;317(7167):1168-9.