Eye-tracking the moving medical image: Development and investigation of a novel investigational tool for CT Colonography

Dr Emma Louise Helbren

MBChB, MRCS, FRCR

University College London (UCL)

Submission for the degree of Doctor of Medicine (Research)

Declaration

I, Emma Louise Helbren, confirm the work presented in this thesis is my own. Where assistance has been received, it has been acknowledged and information which has been derived from other sources has been indicated and referenced appropriately.

Abstract

Colorectal cancer remains the third most common cancer in the UK but the second leading cause of cancer death with >16,000 dying per year. Many advances have been made in recent years in all areas of investigation for colorectal cancer, one of the more notable being the widespread introduction of CT Colonography (CTC).

CTC has rapidly established itself as a cornerstone of diagnosis for colonic neoplasia and much work has been done to standardise and assure quality in practice in both the acquisition and interpretation of the technique. A novel feature of CTC is the presentation of imaging in both traditional 2D and the 'virtual' 3D endoluminal formats. This thesis looks at expanding our understanding of and improving our performance in utilizing the endoluminal 3D view.

We present and develop novel metrics applicable to eye-tracking the moving image, so that the complex dynamic nature of 3D endoluminal fly-through interpretation can be captured. These metrics are then applied to assess the effect of important elements of image interpretation, namely, reader experience, the effect of the use Computer Aided Detection (CAD) and the influence of the expected prevalence of abnormality. We review our findings with reference to the literature of eye tracking within medical imaging.

In the co-registration section we apply our validated computer-assisted registration algorithm to the matching of 3D endoluminal colonic locations between temporally separate datasets, assessing its accuracy as an aid to colonic polyp surveillance with CTC.

Impact Statement

CTC has led the way in incorporating the use of 3D ‘virtual’ imaging into routine clinical radiological practice. This novel format has generated new challenges in understanding and learning which need to be confronted if we are to develop efficient mechanisms for training and quality improvement.

This thesis demonstrates and develops the application of a set of detailed and reproducible metrics to harness the valuable information obtainable through eye-tracking studies in the 3D medical imaging environment. These metrics can be applied to compare reader performance under a range of conditions and therefore have the potential to elicit important information about the way we view and interpret 3D imaging. It is hoped these metrics will be adopted by other groups to analyse reader behaviour in the 3D environment, not just in CTC but also more widely across different forms of medical imaging and possibly beyond.

The application of a computer-assisted registration algorithm developed at our institution to co-register temporally separate CTC datasets aims to demonstrate the potential to extend the applications of such software to different areas of challenging clinical practice. The continual pursuit of ways to enhance our accuracy and efficiency, as is demonstrated in this work, has the potential to deliver sizable gains in service delivery if implemented on a wider scale.

Table of contents

Preface

The research included was undertaken whilst a member of the Centre for Medical Imaging (CMI) at University College London (UCL) and University College London Hospital (UCLH) under the supervision of Professors Steve Halligan and Stuart Taylor.

This work was funded by the UK National Institute for Health Research (NIHR) under its Programme Grants for Applied Research funding scheme (RP-PG-0407-10338). A proportion of this work was undertaken at University College London and University College London Hospital, which receives a proportion of funding from the NIHR Biomedical Research Centre funding scheme.

Funding was also provided by the EPSRC-CRUK Comprehensive Cancer Imaging Centre of UCL and KCL (Grant No. C1519AO).

Views expressed are the authors own and may not be shared by the supervisors or supporting institutions.

Contribution and collaboration statement

As mandated by modern medical research, I have collaborated widely on the studies included here and my role and the role of others is described clearly throughout. Work that is by others is acknowledged and referenced appropriately.

The studies detailed have not previously been submitted as part of a postgraduate thesis.

Understanding reader behavior and performance in CTC through the utilization of eye-tracking

To deliver high quality research in this novel research field, the collaboration of specialists from a range of scientific domains has been essential. Under the supervision of Professor Halligan, I have guided the aims and objectives of each project with a focus on maintaining relevance to the radiological task of CTC interpretation, working to ensure appropriate reader participation, selecting and preparing imaging for use within each study and presenting the findings to the wider radiological community in an accessible form.

To ensure the accusation of high quality eye-tracking data, Dr. Peter Philips of Cumbria University, an image perception scientist with 8 years' experience, oversaw eye-tracking data acquisition and processing as well as assisting with video preparation.

Statistical advice and analysis has been provided by Dr. Susan Mallett and Dr. Thomas Fanshawe. Their expertise has helped us handle the exceptionally large data set generated and enabled us to manage the many

complex elements which the data and reader relationships presented us with.

The early preparatory work of Dr. Darren Boone (Radiologist) who participated in the group prior to my appointment is also gratefully acknowledged. He was instrumental in securing ethical approval for these studies and where we have utilized cases he has prepared or data he has collected, this has been acknowledge.

All aspects of these studies have been rigorously reviewed by all collaborators and a great deal of time and resource has been given by all. The need for this level of commitment, by all members of the research team, to be reflected in the published literature led the group to agree that it was right for us to share authorship of the papers generated by this work. The ordering of authors within these studies is therefore intentionally varied. The contribution of all four of the research workers state above to these studies was considerable regardless of stated order on publication. Without the unique input of each specialist and the supportive, constructive ethic shown within the group, these studies would not have been possible. I am very grateful both to my hardworking collaborators and senior supervisors (Professor Steve Halligan, Professor Stuart Taylor and Professor Doug Altman) who were all highly supportive of this model of working.

Author Declaration – Developing registration software to improve reader diagnostic performance in CTC

Work presented in this section was produced by a research group under the supervision of Professor Steve Halligan and Professor David Hawkes.

The author was the sole medical research fellow within the group for this study and led the publication arising. The author significantly contributed to the study design and data collection, identifying and confirming polyp locations and calculating registration error. I also presented this work as a scientific presentation at the ESGAR conference 2013.

Other group members were computer scientists, Dr. Holger Roth and Thomas Hampshire who had developed the algorithm. Dr. Holger Roth undertook the endo-luminal registration and helped with calculation of registration error.

Cases were provided by Professor Perry Pickhardt who also advised on study design, as did Professor Stuart Taylor.

Acknowledgements

I wish to thank Professor Steve Halligan and Professor Stuart Taylor, whose wisdom, guidance, support and immense patience have enabled me to complete this work.

I have had the great honour to collaborate with a wide range of talented and inspiring teams and individuals and teams during this work, namely;

Dr Tom Franshawe and Professor Doug Altman (Centre of Statistics in Medicine, University of Oxford) and Dr Susan Mallet (now of the Test Evaluation Research Group, Birmingham University).

Dr Peter Phillips (University of Cumbria)

Professor David Hawkes, Dr Holger Roth and Dr Tom Hampshire (Centre for Medical Image Computing, UCL).

Dr Andrew Plumb (UCL) and Dr Darren Boone (University College Hospitals), whose friendship and exceptional work on these topics preceding and following these studies has made this thesis possible.

Thanks also go to Heather Fitzke and Nichola Bell who gave administrative support to this study during their time at the Centre for Medical Imaging, UCL.

Thanks also go to the ESGAR (European Society for GastroIntestinal and Abdominal Radiology) for allowing me to collect data for our eye tracking studies from delegates during their annual conference and CTC workshops.

I am very grateful to Vital Images (Vital Images Inc., Minnetonka, Minnesota, USA) and iCAD (iCAD Inc., Fairborn, Ohio, USA) for provision of the medical image workstation used for this research.

This work was made financially possible through the support of the NIHR (programme grant RP-PG-0407-10338) and I am grateful for the opportunity this had afforded me.

Ethical approval statement

Eye-tracking studies

Research Ethics Committee (REC) approval was obtained to eye-track consenting observers prospectively for studies in Chapters 3, 4, 5 and 6 of this thesis; Project ID: 10/0051. REC reference number: 09/H0304/66. Protocol number: Version 1.2. Study Title: “CT colonography: Effect of Computer-Assisted-Detection on Diagnostic Performance and Visual Search. Granted on 10th September 2009 from the Cambridgeshire 1 Research Ethics Committee (Chair Dr Daryl Rose).

These studies utilised anonymised CTC data from Institutional Review Board (IRB) and Research Ethics Committee (REC) approved studies which had collated data from patients in both the USA and Europe [1, 2].

Co-Registration

Anonymised CTC data collected from asymptomatic individuals undergoing screening for colorectal polyps in the USA was utilised for the study included within Chapter 7. Our commercial partner, Medicsight PLC, confirmed that all ethical approvals were in place to use this data, namely Institutional Review Board (IRB) and Health Insurance Portability and Accountability Act (HIPAA) approvals.

List of Tables

Table of Figures

Glossary

2D	Two dimensional
3D	Three dimensional
ACPGBI	Association of Coloproctology Great Britain and Ireland
BCSP	Bowel Cancer Screening Programme
BSG	British Society of Gastoenterology
CI	Confidence Interval
CRC	Colorectal Cancer
CTC	Computerised Tomographic Colonography
ESGAR	European Society of Gastrointestinal and Abdominal Radiology
FIT	Faecal Immunochemical Test
gFOBT	guaiac Faecal Occult Blood Test
IQR	Interquartile range
PPV	Positive Predictive Value
ROI	Region of Interest
SIGGAR	Special Interest Group in Gastrointestinal and Abdominal Radiology
VC	Virtual Colonoscopy

Thesis strategy

UCL is a world leader in the imaging of colorectal cancer and the development of CTC. Professor Halligan, Director of the UCL Centre for Medical Imaging, leads a large and experienced multidisciplinary group working on the diagnostic accuracy and evidence-based implementation of CTC. Professor Halligan was the Chief Investigator for the HTA funded RCT of the technology (HTA 02/02/01; the SIGGAR trial) and held an NIHR Programme Grant for Applied Research in the area. The initial phases of the Programme generated many research questions related to how CTC can be applied and implemented in NHS practice and regarding what further developments can be made to maximise the diagnostic performance of CTC. This thesis aims to contribute further to this knowledge and the further development and refinement of CTC.

Section A provides an introduction to this work with background on colorectal cancer and the range of tests used for its investigation. It then discusses further the technique of CTC. An introduction to eye tracking in medical imaging is then included with reference to previous studies. Finally, we outline the role for co-registration, the development of our algorithm for matching endo-luminal polyp locations and the prior studies validating this.

In **Section B** we present our work on eye tracking within 3D CTC. We present the development of novel metrics for the description of eye tracking in the 3D environment. We then apply these to the study of the characteristics of experienced and inexperienced readers, the study of the

effect of CAD markers on reader gaze and the potential impact of perceived levels of prevalence abnormality in endo-luminal CTC.

Section C presents the application of our computer-assisted registration algorithm to temporally separate CTC studies. We assess the accuracy of the algorithm for predicting polyp location between studies using both a consistency and longitudinal method of analysis.

Finally, with reference to the literature we summarise and conclude our work in **Section D**. This is followed by the appendices and references associated with the manuscript.

Section A: Introduction

Chapter 1: Colorectal cancer and its investigation

Colorectal cancer, frequently known as bowel cancer, is the second commonest cause of cancer death in the UK. Overall five-year relative survival of colorectal cancer patients in England is 50.7%. Five-year survival rates range from 93.2 % for early cancers confined to the bowel wall to 6.6% for those with metastatic disease [3]. Currently only 13% of patients with colorectal cancer (CRC) are diagnosed with early, bowel confined, disease [3].

Most colorectal cancers likely develop from pre-existing adenomatous polyps and therefore early detection and excision of these lesions can reduce the rate of subsequent CRC [4].

The rate of synchronous CRC is 3.9% and a third of these lie within a different segment of the colon [5]. These synchronous tumours are usually not palpable at operation making whole-colon preoperative assessment of the colon essential [6].

Tests for colorectal cancer

A range of tests are now available for the investigation of patients with suspected colorectal cancer, both symptomatic and asymptomatic. The nature of the test used will depend on the patient, their indication/ symptoms, the patient's co-morbidities and wishes, and health care resources.

Non-invasive tests

The guaiac faecal occult blood test (gFOBT) detects haemoglobin via presence of a peroxidase reaction. It has been the primary screening test used by the NHS Bowel Cancer Screening Programme (BCSP) since the programmes introduction in 2006. Stool test kits are posted to residents homes between the ages of 60 and 75 years old every two years and then these are returned through the post for analysis. A positive gFOBT raises the possibility of cancer and therefore leads to an invitation for colonoscopy. This screening method has met with good results with the randomised controlled screening trials demonstrating a 15 % reduction in colorectal cancer related mortality within the screened population, although it is noted, no change in the all-cause mortality [7]. However, gFOBT is not considered of adequately sensitivity for investigation of symptomatic patients; less than 2% of screening tests returning positive [8]. Transition is now being made within the screening programme to replace gFOBT with the Faecal Immunochemical Test (FIT), as the initial screening test. FIT uses antibodies to detect the globulin portion of human haemoglobin. It is therefore specific for human blood, whereas gFOBT is not. FIT is more sensitive than gFOBT and, importantly, its output is continuous so that a threshold can be applied, varying sensitivity and specificity to the needs of the programme. Both gFOBT and FIT are relatively low cost tests but better participation rates have been seen with FIT and the automated analytical processing of FIT over gFOBT minimises error [9]. Data from 19 studies showed the overall sensitivity of FIT for CRC was 0.79 (95% CI:0.69 to 0.86) with overall

specificity of 0.94 (95% CI:0.92 to 0.95) [10]. Again, this is insufficient for symptomatic investigation.

An alternative test is the faecal based DNA test. This has a higher sensitivity than FIT but currently is very expensive (cost $599 per test) precluding its use in national screening programmes.

Blood-based DNA testing (the SEPT9 assay) is still relatively novel and further development is needed if it is to detect adenomas and improve sensitivity for colorectal cancer, which currently stands at 48.2% [11, 12]. Blood based tests have been shown to have a high patient acceptance level though and this approach is therefore likely to remain a target for further development [13].

CEA (carcinoembryonic antigen) is a tumour marker commonly used for monitoring of colorectal cancer patients post surgical resection. Although CEA can be useful to detect cancer recurrence in this group, especially in the presence of multiple metastases (with a 75% sensitivity level) it has a poor sensitivity for the detection of primary CRC with only a sensitivity of 41.7% even in stage III disease [14]. It is therefore inappropriate as a useful diagnostic or screening tool.

Colonoscopy

Colonoscopy is an invasive procedure where a flexible telescopic camera is inserted into the large bowel via the anus. It remains the gold standard reference test for colonic assessment as it allows both direct visualisation of the endoluminal surface of the intestine as well as offering the possibility of

simultaneous biopsy / excision of mucosal lesions. Colonoscopy requires full large bowel purgation and usually the administration of sedatives. Its use extends to all large bowel pathology, it being notably superior to CTC for the diagnosis of colitis. The one exception to this rule is in the diagnosis of diverticulosis, which is better demonstrated on CTC [15]. Colonoscopy is a highly sensitive test which is well established and well-regulated with a network of UK centres and operators examining both symptomatic and asymptomatic screening groups [15-17]. UK endoscopists are quality assured through the Joint Advisory Group (JAG) accreditation system. The BCSP also maintains a database of all colonoscopies undertaken within the programme.

However, colonoscopy is an invasive, time-consuming, and often unpleasant procedure. It also has relatively high morbidity and even mortality, predominantly related to the sedation used. To enable the scope to be passed around the bowel and to visualise the endoluminal surface with clarity, cathartic bowel cleansing with low fibre dietary preparation is required to clean and dry the colon. Because colonoscopy is frequently very uncomfortable it is standard practice to administer intravenous pain relief and / or sedation during the procedure. Serious complications from colonoscopy are rare and are predominantly related to bleeding post-polypectomy, colonic perforation and /or the sedation/ pain relief administered. Metrics for all these complications vary widely, likely reflecting variations in practice. The stated BCSP perforation rate is 0.09%, while the bleeding rate is 0.59%, but with only 0.13% requiring intervention and 0.04% requiring transfusion [17]. Quality assurance within the BCSP is tightly controlled so that complication

rates outside the programme are likely to be higher. Mortality from colonoscopy is rare but does occur in approximately 0.007% [18].

A further problem encountered in colonoscopy is an occasional inability to complete the test because colonic coverage is incomplete ("total colonoscopy" is defined by caecal intubation). The rate of incomplete colonoscopy varies between 10% and 15% [19]. For accreditation within the BCSP a screening endoscopist must maintain a caecal intubation rate of 90% or more. In the SIGGAR trial (of symptomatic patients) 11% of colonoscopies were incomplete with 7% having another test due to this [15]. Reasons for failure include patient intolerance, inadequate bowel preparation, difficult anatomy and the presence of other pathology that obstructs the scope passage (such as diverticular disease).

Flexible sigmoidoscopy

Flexible sigmoidoscopy (FS) is an endoscopy using a shorter scope that is limited to the colo-rectum distal to the splenic flexure. Flexible sigmoidoscopy requires less preparation (only the left colon need be emptied), is quicker, and requires less resource than a full colonoscopy and therefore can be appealing to patients and health institutions alike. Further, the left colon is less prone to perforation than the right. As two-thirds of colorectal neoplasms are found in the sigmoid and rectum flexible sigmoidoscopy is also often sufficient to identify the majority of neoplastic pathology [20]. Patients are usual prepared only with an enema on the day of the test, in contrast to the full cathartic and dietary preparation required for colonoscopy.

Flexible sigmoidoscopy has been shown to be a safe and clinically effective investigation for colorectal cancer in patients with change in bowel habit to looser and /or more frequent stool alone or in patients with rectal bleeding, without anaemia and/ or an abdominal mass [20], i.e. patients presenting with “left-sided” symptoms.

Large randomised controlled European trials have shown significantly reduced mortality for all stages of colorectal cancer in those undergoing flexible sigmoidoscopy screening. The English BCSP introduced one-off flexible sigmoidoscopy at 55 years old in selected centres in 2013, to run prior to FOBT screening. Patients with high risk findings on flexible sigmoidoscopy proceed to completion colonoscopy. This initiative was known as ’Bowelscope’ and role out to the whole of England is planned [21].

The limitation of flexible sigmoidoscopy is that due to lack of total bowel cleansing and the standard practice of not giving aggressive sedation or intravenous pain relief, the depth of intubation achieved can be limited by residual stool and/ or pain. The lack of a clearly definable endpoint during the test (such as caecal intubation for colonoscopy) means that completion of the test to the splenic flexure is highly subjective, leading to variability in effectiveness in clinical practice. In reality a minority of scopes actually enter the descending colon and a quarter even fail to complete examination of the sigmoid [22].

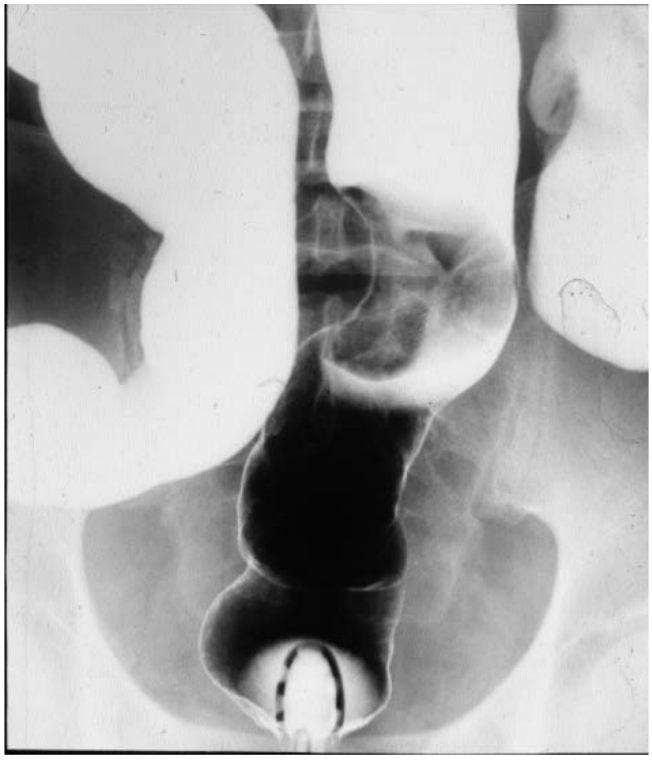
CTC and imaging of the colon

Barium enema

For decades the mainstay of colonic imaging was the barium enema. This double contrast study utilised cathartic colon cleansing followed by barium suspension instilled per rectum. Gas insufflation via a rectal tube was then performed to distend the large bowel which had been coated by the barium suspension. Images were obtained via fluoroscopy (Figure 1).

Figure 1: Barium enema

Double contrast image (barium and air), showing a cancer in the recto-sigmoid.



To ensure the bowel wall was visualised in its entirety without overlap, images were obtained by repositioning the patient on the fluoroscopy table at multiple points during the study.

Limitations of barium enema included patients' ability to retain both the bowel contrast and gas, the ability to adequately coat the entire colon with contrast, and difficulties in interpretation. The fluoroscopic nature of image acquisition also meant that only intraluminal abnormality could be assessed.

The advent of CT, initially as an unprepared and undistended study and later as CT Colonography (CTC) with cathartic preparation and automated carbon dioxide insufflation, lead to the gradual demise of the barium enema.

The publication of the multi-centre randomised SIGGAR trial in 2013 confirmed the significantly inferior performance of barium enema when compared to CTC for detection of polyps >1cm and cancers in symptomatic patients [23]. Lower detection rates by barium enema have also been demonstrated within screening populations [24].

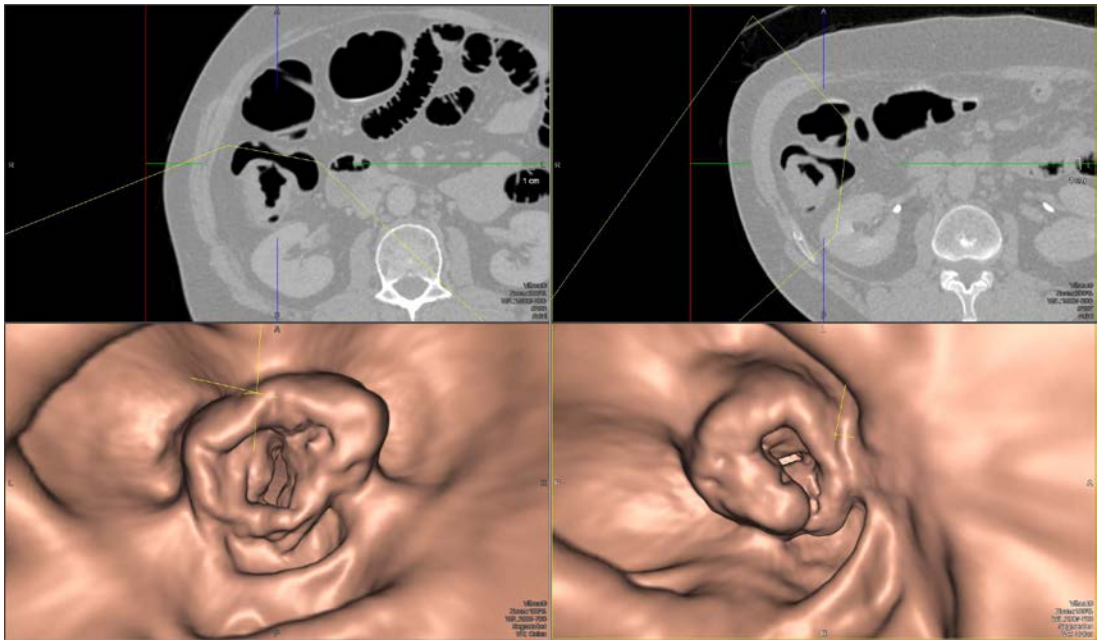
CTC is now the recommended radiological imaging technique for diagnosis of colorectal cancer and premalignant lesions by all major European and North American associations and within the UK Bowel Cancer Screening programme. [23, 25]

CTC

CTC, also known as Virtual Colonoscopy (VC) is a method of examining the large bowel that presents images to the interpreting radiologist in both traditional 2D and more novel 3D visualisation formats (Figure 2).

Figure 2: CTC

Screen shot from a CTC study showing a cancer in the ascending colon. 2D and 3D images are presented from acquisitions with the patient in the supine and prone positions.



Introduced in the mid 1990's, CTC is a relatively new technique that has quickly replaced barium enema and offers a competitive alternative to the gold standard test for colorectal cancer, the colonoscopy.

The first ESGAR (European Society of Gastrointestinal and Abdominal Radiology) consensus statement on the use of CTC came in 2007[26] and other international guidelines soon followed [27, 28]. A further second ESGAR consensus was published in 2013 [29]. CTC has generated considerable interest in both medical and lay arenas. It has the ability to identify colonic cancer and its precursor, the adenomatous polyp, in a less invasive manner than endoscopy, allowing a safer method of investigation for

more frail, high-risk patient groups. There has been extensive work published on all facets of the technique, with the list of indexed publications now approaching 4000.

When CTC is combined with intravenous contrast and a CT Chest examination, the complete radiological diagnosis and staging of colonic cancer is achievable in a single examination. It is also able to delineate and characterise other colonic features clearly, i.e. strictures, diverticula and post-surgical anatomy. Extra-colonic pathologies both benign and malignant can also be identified and assessed, e.g. renal lesions and abdominal aneurysm.

The technically impressive virtual 3D endoluminal images from inside the colon have captured the imagination of many, both medical and lay-persons. The ability to ‘fly through’ the bowel (viewing the endoluminal surface of the colon as a video with fixed navigational speed) represents a paradigm shift in imaging and data interpretation and asks new questions of how medical imaging is presented, viewed, and interpreted.

Colonic preparation and faecal tagging

Patient preparation is required prior to CTC for the assessment of polyps and cancer in both symptomatic and asymptomatic groups. Regimes vary between institutions but the traditional components have been dietary restriction via a low fibre diet, bowel purgation with laxative agents, and oral positive contrast for tagging of any residual stool and/or fluid.

Diet

Use of a low fibre diet reduces faecal volume leaving significantly less untagged faeces and reducing stool heterogeneity [30]. It is standard practice to provide written information to patients describing the local preparation schedule included within a ‘diet sheet’ detailing foods to avoid and those advised prior to CTC; low fibre foods include white bread, rice and pasta. Lengths of dietary preparation vary but are generally for one to two days prior to the day of investigation.

Laxatives

Laxative agents vary in the aggressiveness of their catharsis. Agents described within the literature include sodium phosphate, magnesium citrate, and polyethylene glycol but the most common used in UK practice is the first, sodium phosphate or picosulfate (brand name Picolax). Although effective for purgation, laxative agents can cause electrolyte disturbances and nephropathy. Often the reason for selecting CTC over colonoscopy is that the clinician wishes to avoid harsh purgative catharsis. Therefore, there is an increasing trend towards the reduction/ elimination of a specific cathartic agent in the preparation for CTC, instead exploiting the laxative effect of tagging agents to produce a more gentle cleansing effect whilst also marking residual stool.

Tagging

Tagging residual faeces within the colon facilitates easier differentiation of stool from polyps or mass lesions. This improves specificity and also reported levels of patient comfort [31]. The greater confidence faecal tagging

gives also potentially reduces the referral rate for subsequently negative colonoscopy [15, 32]. Hyperosmolar iodinated contrast agents, such as diatrizoate meglumine (Gastrografin) have a mild to moderate laxative effect that is often sufficient to achieve an appropriately clean bowel for CTC, allowing elimination of sodium picosulfate. Regimens with diatrizoate meglumine can deliver high PPV studies, even with a low dose protocol [33] and are commonly acceptable to patients [34]. Other regimens with low cathartic preparation, just as with barium-based tagging regimens, have also shown good results [35, 36] but barium tends to be a less desirable agent, producing a more heterogeneous tagging of stool and fluid [37]. Iodine based preparations which give residue stool a more homogenous density are desirable. Barium and iodine combined stool and fluid tagging necessitates a more complex preparation scheme. This may then reduce patient compliance.

As an adjunct to these cleansing and tagging regimens, electronic (computer generated) cleansing applications are also available. These may improve polyp conspicuity [37] [38] but are not always available within standard software packages. They also varying in efficiencies between manufacturers, with some effectively performing a “digital polypectomy” for small lesions.

Colonic insufflation

Colonic distension with an automated carbon dioxide insufflator is the method of choice (over manual insufflation with room air or carbon dioxide) to achieve best luminal distension [39]. Although patient discomfort may be increased during the procedure with automated CO2 insufflation, this

subsides quicker than using the air alternative [40] as carbon dioxide is reabsorbed approximately twenty times faster. Insufflation is achieved by insertion of a thin flexible tube into the rectum by an appropriately trained health care professional. Inflation of a small balloon at the tip with air is optional to help gas retention, but deflation is advisable on one acquisition to prevent compressing and thereby obscuring, low lesions. Hyoscine-N-butylbromide (buscopan) is the spasmolytic of choice [41, 42] to aid adequate colonic distension but must be used with caution in patients at risk of side effects, specifically those with cardiac disease when administration can prove fatal [43]; the agent causes tachycardia.

Colonic distension for CTC is only ‘sufficient’ once all segments of the colon can be visualised in at least one patient position. The volume of gas delivered should not be used as a surrogate of distension adequacy, rather appearances of the scout image balanced against patient comfort should be used to judge when sufficient distension is achieved. A scout image should be reviewed before each acquisition, i.e. in each patient position prior to scan to ensure this has been achieved, prior to acquiring the data.

Once the examination is complete, the study should be reviewed by a trained practitioner to assess for technical adequacy and (rarely) possible perforation.

A systematic review of colonic perforation during screening CTC quoted the rate as <2 in 10,000 procedures [11], while another group reported no perforations in 11707 screening CTCs and only 2 in 10216 symptomatic studies [44].

Factors increasing the risk of perforation during CTC include an obstructing cancer, severe diverticulitis or colitis, hernia, and recent polypectomy / biopsy. Perforation is more likely with older age and concomitant colonic disease [45]. CTC following same day polypectomy though is not infrequently performed and there is growing evidence that initial concerns in this setting may have been exaggerated [46]. If there has been difficult optical colonoscopy directly prior to CTC, then a low dose CT should be considered prior to CTC to ensure no perforation has occurred. If appearances of this are satisfactory, it is appropriate to continue with a standard CTC procedure, including use of a colonic automated insufflator [47]. If concern persists then it may still be prudent to delay CTC, although the precise interval is unclear[29].

Image acquisition

Image acquisition for CTC is by multi-detector row scanners obtaining CT through the whole abdomen and pelvis in a single breath hold. Maximum collimation should be no more than 2.5mm although thinner slices are preferable [29]. Reconstruction with a 20 to 30% overlap is used.

Supine and prone acquisition is standard but if the patient is unable to lie prone then lateral decubitus acquisitions may be obtained instead[48]. Dual position acquisition aids differentiation of mobile stool and faecal tagging solution from endoluminal polyps /cancers, the latter remaining fixed to the mucosal surface with the former being mobile, preferentially lying in dependent portions of bowel. Identifying the same polyp in both acquisitions

in a fixed position enhances diagnostic confidence when reporting CTC studies.

In approximately 10% of routine CTC studies it will also be required to complete a third acquisition to ensure adequate distension and assessment of the colon is complete [49].

Within the UK BCSP, screening populations are scanned using a low dose protocol with no IV contrast administered so as to reduce the potential for adverse events. If a likely malignant lesion is seen on the first acquisition there is always then the opportunity to obtain the second acquisition as a full staging study, with supplemental CT chest and IV contrast. A low dose study is considered one in which the median effective dose is less than 5.7mSv [50]. 120kV should be used for both supine and prone acquisitions. Less than or equal to 50mAs is preferable (weight dependent). It is difficult to be too prescriptive regarding the radiation parameters for scanning as dose will be affected by patient size and the scanner used, but protocol is key to obtaining a diagnostic image at a dose which is as low as reasonably possible [51].

In symptomatic patients, without known colorectal cancer, use of IV contrast is dependent on the scan indication and the need to evaluate extra-colonic structures. When IV contrast is utilised, it is generally administered for the supine study alone. Whether this is preceded by a low dose, non-contrast, prone acquisition or followed by a low dose, prone, delayed phase acquisition is dependent on local protocol but the low dose acquisition should be less than or equal to 50mAs. Inevitably, the portovenous phase scan will

require a higher radiation dose than the other acquisitions and this is acceptable when balanced against clinical requirements. There is arguable benefit for first acquiring a non-contrast prone acquisition and then, if there is a probable colonic tumour, following with a standard supine chest, abdomen, and pelvis staging study with intravenous contrast.

Unless contraindicated, intravenous (IV) contrast should be administered to all patients with known colorectal cancer, so as to facilitate more accurate staging [52]. If available, dose modulation and iterative reconstruction should be applied which, within appropriate parameters, will also dose reduction without loss of image quality. [53].

Reading paradigms

CTC is generally interpreted using a combination of two different visualisation methods; the 2D ‘standard’ axial reading format akin to other CT examinations and the 3D ‘Virtual colonoscopy’ or ‘fly-through’ view. Both CT acquisitions should be interpreted on an appropriate medical image display workstation that provides these visualisations. Many commercially available workstations have a range of different 3D visualisation formats, i.e. ‘virtual dissection’ / ‘filet’ / ‘panoramic view’, and it is important that readers are fully trained in these techniques with an understanding of the benefits and limitations of both their specific visualisations and, more globally, the 2 and 3D formats in general.

The primary read may use either the 2D visualisation with subsequent 3D review used for problem solving, or use a primary 3D read with 2D for problem solving and review of extra-colonic findings. Although the decision

regarding whether to adopt a primary 2D or 3D approach lies with the reporter (and available software), there is evidence that lesion sensitivity is improved when adopting a primary 3D approach [54, 55]. Specificity, though, is aided by 2D review [56] and it is possible that the improved results achieved with primary 3D read may be more a reflection of a surrogate for workstation quality, training and time available for reporting [24]. Preference for the 2D read may reflect the ability to assess both intra and extra colonic findings, aiming for a more global assessment of pathology with less focus on smaller intra-colonic findings. Observers new or inexperienced at reading CTC, will generally find the 2D reporting format more familiar and it will also be faster than the more specialised 3D format, again suggesting that a 2D preference may reflect inexperience rather than a measure of intrinsic visualisation deficiency.

Colonic lesions should be measured in the plane and on the visualisation on which they are best demonstrated. A polyp stalk, if present, is excluded from any diameter measurement. The segmental location of the lesion should be reported. Extra colonic regions are assessed as for standard CT reporting, noting limitations if the acquisitions are low dose and / or unenhanced.

Increasing radiologist experience with CTC is associated with higher detection rates of abnormality and higher PPV. The number of individual cases required to demonstrate such an improvement are vast with Plumb et al suggesting that more than 1000 studies are required to benefit from such an effect within the BCSP given the low prevalence of abnormality in screening populations [55]. Currently the Royal College of Radiologists requests radiologists participating in the BCSP to have completed a formal

training course and to have read 50 CTC studies validated by colonoscopic verification, maintain a practice of reading at least 100 CTC cases per year, and to participate in colorectal MDT activities and service audit. Quality assurance of CTC within the BCSP includes auditing of a number of different facets of patient safety, outcome and experience, aimed at improving standards [57, 58]. Participation in formal training courses has been shown to improve reporting accuracy [59]. Many, however, only attend training courses once they are already reporting CTC [60].

CAD: Computer Aided Detection

Computer aided detection (CAD) is an adjunct to CTC interpretation available on some image display workstations. CAD aims to highlight potential colonic polyps to the reader. In general, CAD does not aim to identify larger cancerous lesions, although this may be achieved serendipitously since these often exhibit polypoid features.

CAD maybe used either concurrently alongside CTC review or after a conventional, unassisted review, as a so called ‘second reader’. Acceptability depends on the licensing stipulations with respect to the software. Both strategies can achieve similar sensitives for detection of polyps >6mm. The second read paradigm, by definition, takes longer to perform, while the concurrent read may sacrifice sensitivity for small (<6mm) polyps [61]. This said, the increase in time needed for the second read paradigm is relatively small and the increased sensitivity achieved does not trade off specificity when considering polyps 6 to 9mm [62]. The benefit of using CAD is more marked, and arguably confined to, inexperienced readers [63]. Limitations

can include a high number of false positive CAD prompts generated in poorly prepared studies and the inability to detect lesions covered by faecal residue. It is also important to remember that the final decision to report a lesion rests with the observer and therefore CAD positive lesions can be dismissed incorrectly.

Lesion size and morphology

CTC aims to identify colonic tumours and polyps. It Is important to examine the attenuation of any lesion detected to ensure they are indeed soft tissue and don't represent foci of faeces, oral tagging or fat attenuation lipomas.

When a polypoid lesion is identified, it can be described on the basis of three morphological subtypes: sessile (broad based), pedunculated (with separate stalk) and flat (elevation above the surrounding mucosa of 3mm or less).

Most colorectal cancers arise via the adenoma-carcinoma sequence; epithelium proliferates to form adenomas, adenomas develop dysplasia which then in turn develops into frank carcinoma. The precursor lesions in this sequence, which must be identified and removed to prevent subsequent carcinoma, are adenomatous polyps. Although geographic variation is present between the East and West, the adenoma-carcinoma sequence is estimated to account for around 83% of colonic carcinomas in total with a relatively high proportion of left sided colonic cancers developing via this route than compared to right sided malignancies [64].

An alternative, though less common, route for the development of colorectal cancer is via the sessile serrated polyp. Sessile serrated polyps may be of

any morphological subtype but are most commonly flat. When >3cm in size they are referred to as ‘carpet’ lesions. These flat lesions are generally right-sided and challenging to identify for both radiologists and colonoscopists alike. Although some success has been reported in expert centres regarding detection of flat lesions, it is generally accepted that CTC is less sensitive for flat lesions than other polyp morphologies [65-67].

The likelihood of dysplasia and cancer within a lesion increases with lesion diameter and patient risk factors, namely patient age [68]. Statistics from large polypectomy datasets of asymptomatic individuals found 1.7% of polyps under 5mm demonstrated advanced histology but none represented carcinoma, increasing at 5 to 10mm to 10% with advanced histology and 0.9% with carcinoma [69]. Above 1cm the carcinoma rate in polyps of 1 to 2cm rises to 2.41% and at >2cm rises to 19.35% [70]. Styker et al demonstrated that polyps of >1cm identified on barium enema but left unresected demonstrated malignant transformation rates of 2.5% at 5 years, 8% at 10 years and 24% at 20 years [71].

Relatively little significant disagreement is observed between CTC and endoscopic measurements with neither being able to claim definitive accuracy [72]. When lesions are measured, ESGAR guidelines suggest that narrow window levels should be avoided to improve measurement accuracy.

Once a lesion is identified the decision then arises as to whether to resect it or to survey. Resection is not without risk and, of course, cost, and if patient co-morbidities exist then these must also be taken into consideration. As

above, most colonic lesions are not cancer and, even if this does develop, development is often slow.

Polyps of 6 to 9mm generally remain stable and some can actually regress over time [73]. So, although all polyps or flat lesions of 6mm or larger should always be reported, the option remains for them to be surveyed, either by CTC or colonoscopy, or to progress to endoscopic resection [29, 73, 74].

Following adenoma removal by colonoscopy, either as the primary investigation or following CTC, the patient then enters a surveillance group stratified by perceived risk. British surveillance guidelines are defined by the British Society of Gastroenterology (BSG) and Association of Coloproctology of Great Britain and Ireland (ACPGBI) and aim to survey low risk patients (1 to 2 adenomas, both <1cm) at 5 years or not at all, intermediate risk patients (3 to 4 adenomas or at least one >1cm) at 3 years, and high risk patients (5 or more adenomas or 3 or more adenomas with one >3cm) at 1 year [75].

CTC: Diagnostic performance in symptomatic patients

CTC and colonoscopy have comparable detection rates for CRC and large polyps in symptomatic patients. The SIGGAR trial demonstrated no significant difference in detection rates between CTC and colonoscopy for large polyps (1cm or greater) and for cancers. A systematic review and meta-analysis by Pickhardt et al reported similarly comparable results with sensitivity of CTC for colorectal cancer of 96%, comparable with colonoscopy [76]. When examining these figures, it is also important to accept that although colonoscopy remains the gold standard of investigation, it too has

an inherent false negative adenoma detection rate, which in tandem colonoscopy studies has been shown to be as high as 36% [77].

Comparative studies point to a slight patient preference for CTC, which is considered more acceptable than colonoscopy, causing less physical discomfort and psychological concern [78, 79]. It is known that colonoscopy precipitated by false positive CTC takes significantly longer than for other indications since the endoscopist is searching for a lesion that is not actually present.

In failed colonoscopy due to a stenotic lesion, preoperative CTC has a NPV of 100% for synchronous cancer and 97% for advanced neoplasia [80].

CTC: Diagnostic performance in screening populations

When screening asymptomatic individuals, CTC was initially hailed as achieving similar detection rates to endoscopy and therefore potentially offering an alternative primary screening test to colonoscopy. [81]. Meta-analyses have shown equivalent sensitivities can be achieved for CTC when compared to colonoscopy when CTC is performed in experienced centres, Johnson et al showing that 90% of adenomas or cancers measuring 10mm or more were identified by CTC [82]. In reality however, detection rates with CTC in more generalised screening centres are likely to be much lower: The English Bowel Cancer Screening Programme (BCSP) reported significantly lower detection rates of both cancers and large (>1cm) colonic polyps when colonoscopy and CTC were compared; 9% v 4.5% and 20.6% v 12.4% respectively [24].

Due to the high prevalence of clinically relevant lesions in FOBT/ FIT positive individuals, it is also questionable whether CTC (which has no therapeutic value) makes sense within the screening programme [82-85]. The fact, though, remains that colonoscopy is not without risk and if a high risk patient is identified as positive on FOBT/FIT via the screening programme and requests / requires further investigation, CTC may provide a less intrusive method to identify or exclude a significant lesion.

C-RADS

Zalis et al proposed a standardised reporting template, “C-RADS”, for findings on CTC[74]. C-RADS is analogous to BI-RADS used for breast cancer screening, and was developed by the Working Group on Virtual Colonoscopy. The rationale was three-fold: to standardise reporting for clarity of clinician and patient; to allow a unifying nomenclature that would allow comparison across differing centres; and to facilitate quality assurance.

C-RADS has since been adopted by the BCSP with a summary grade given on each screening CTC report (as outlined in Table 1).

Table 1: Grading of colonic findings

C0. Inadequate Study/ Awaiting prior comparisons

- Inadequate prep: cannot exclude lesions >or equal to 10mm owing to presence of fluid/faeces.
- Inadequate insufflation: on or more colonic segments collapsed on both views
- Awaiting prior colonic studies for comparison

C1. Normal Colon or Benign Lesion: Continue Routine Screening (1)

- No visible abnormalities of the colon
- No polyp > or equal to 6mm
- Lipoma or inverted diverticulum
- Non neoplastic findings e.g. colonic diverticula

C2. Intermediate Polyp or Indeterminate Finding: Surveillance or Colonoscopy Recommended (2)

- Intermediate polyp 6 to 9mm, <3 in number
- Indeterminate findings, cannot exclude polyp >or equal to 6mm in technically adequate exam

C3. Polyp, Possibly Advanced Adenoma: Follow-up Colonoscopy Recommended

- Polyp > or equal to 10mm
- >or equal to 3 polyps, each 6 to 9mm

C4. Colonic Mass, Likely Malignant: Surgical Consultation Recommended (3)

- Lesion compromises bowel lumen, demonstrates extracolonic invasion

1: Every 5 to 10years

2: Evidence suggests surveillance can be delayed at least 3 years, subject to individual patient circumstances

3: Communicate to referrer. Depending no local practice, endoscopic biopsy may be indicated

Extra-colonic findings

The C-RADS classification also addresses extracolonic findings to help both simplify and convey the significance of abnormalities identified beyond the colon (Table 2).

Table 2: Grading of extra colonic findings

E0. Limited Exam. Compromised by artefact; evaluation of extracolonic soft tissues is severely limited.
E1. Normal Exam or Anatomic Variant. No extracolonic abnormalities visible.
E2. Clinically Unimportant Finding. No work-up indicated.
E3. Likely Unimportant Finding, Incompletely Characterized.
E4. Potentially Important Finding.

In the SIGGAR trials, 58.6% of patients having CTC had extra colonic abnormalities reported. Most were unimportant, with only 8.2% going on to have an additional investigation specifically for these findings. The extra-colonic cancer rate was ultimately 1.6% [86]. Other studies have reported similar rates for identification of extra colonic malignancy, for example, 0.35% in screening populations [87] and 1.9% within symptomatic groups [15, 23]. However, it is important to also consider that not all significant extra-colonic pathology will be malignant, for example, size significant abdominal aortic aneurysm. Returning to the SIGGAR trials, 35.6% who underwent additional investigation received an extracolonic diagnosis that explained their presenting symptoms [86].

Opinion remains divided between those who consider diagnosis of extra-colonic pathology, especially cancer, as a considerable asset for CTC and those who point to the downstream risk, cost and anxiety that such findings precipitate, often proving eventually to be CTC false positive findings. Of course, patient care revolves around diagnosis and explanation for their symptoms which frequently cannot be resolved by a single test. The eventual diagnostic outcome is often unchanged, irrespective of the diagnostic route taken [15].

Chapter 2: Novel techniques in CTC; Eye-tracking and visual perception in imaging research

Eye tracking allows assessment of gaze and uses this information to gain a greater understanding of observer behaviour in visual perception tasks. An eye tracker unit assesses eye movement by projecting near-infrared light onto the eyes and uses cameras to capture high frame rate images of the reflection pattern off the eye, an example of an eye tracker unit (as used in this thesis) is showed in Figure 3. The information gathered is then processed to obtain data on eye position and gaze point, which can be used to either monitor behaviour or enable hands-free interaction with screens and devices.

Figure 3: Eye-tracking unit set-up

Eye-tracking unit set-up as used in this thesis. The unit sits below the computer display screen, being unobtrusive to the reader.

Eye tracking allows the recording of where gaze rests on an image display and then, by inference, exactly where the observer's attention is drawn. Eye

tracking is used in both research and commercial development across sport, commerce, and medicine, including medical imaging, to try and gain a greater understanding of observer search, focus, and decision making.

The centre of our field of view gives sharp detail to our vision [88, 89]. Pausing the centre point of our gaze on a feature of interest is called ‘foveal fixation’. Our vision moves between these points of interest via rapid jumps called ‘saccades’, during which time no visual information is processed. Returning repeatedly to the same point of interest is referred to as a ‘spatial clustering’[90].

Eye tracking has long been used to study observation of traditional 2D images of pathology such as lung lesions [91], breast lesions [92] and bone fractures [93]. It has attempted to characterise effective search characteristics [94] and explain the reason for radiological error [95, 96]. It has also been used to assess image display formats [97, 98], reading techniques [99] and even the required characteristics for second human readers [100].

Traditional imaging paradigms are static and 2D, the most obvious being an x-ray. More modern medical image display may utilise representations of 3D data and also images that are moving, such as in CTC. The challenge for eye-tracking such images is that fixations are replaced by smooth pursuit eye movements with gaze fixed on structures as they move across the display screen. Not only does gaze then become a dynamic process but the visual stimuli are also dynamic in position, size and contrast.

Expertise and experience

There have been multiple and varied attempts to use information obtained from eye tracking analysis to improve the diagnostic performance of readers [101]. This has led to the studies that investigate whether visual search patterns differ significantly in readers of differing levels of experience. For mammography it has been possible to demonstrate a significant difference in readers' gaze duration, scan paths, and detection times based on their level of experience [92]. Other successful studies include that by Hu et al who found that more systematic scanning patterns were observed with experienced versus inexperienced readers when searching for bone fractures [93]. Kundel et al identified an improved global perception in experts, with the ability to pinpoint abnormality without a 'search-to-find' strategy [94, 102-104]. In another study, Kundel argued this effect went further, and that experts interpreting chest radiographs were able to correctly identify pathology using peripheral vision alone, without fixation [91].

The message from the literature, though, is complex and often not unified. Hu et al highlighted, for example, that different forms of image resulted in different patterns of visual search, i.e. inter-cluster jump distances were greater for chest images than for bone. Positive decisions were associated with prolonged gaze durations but prolonged gaze was also significantly longer in false-negative versus true-negative reader decisions [93]. Barrett et al also found that differing forms of pathology in mammography where associated with differing search characteristics, for example, long dwell times were associated with an increased false-positive diagnosis of mass

lesions whereas for micro-calcifications longer dwell times were associated with a higher true-positive outcome [105].

There has always been some difficulty unifying the descriptors used for eye tracking research of medical images. Large amounts of raw data are generated by eye tracking and, as the nature of visual behaviour is poorly understood, comparing data using scientific, reproducible, and valid methodologies can be challenging [106]. The varying complexities of the imaging paradigm being studied also frequently necessitates differing methods of data analysis. This has led to often complex analytical techniques being created principally for the task in hand[107-109]. The range of hardware, settings, and environmental conditions within viewing environments also introduce additional possible sources of variability and precipitate error when research findings are generalised to wider practice [110].

As a consequence, it is unsurprising that the simple three categories of error described by Kundel in 1978 continue to influence image perception research today: the error of search or scanning, (when the pathology is never fixated/ there is an absence of gaze pursuit and/or fixation on an ROI); the error of recognition (when fixation happens for a short period but a lesion is not recognised as important); the error of decision (where fixation occurs but the reader fails to categorise the abnormality correctly after scrutinising the pathology)[95].

Computer aided detection (CAD)

Eye tracking has aided our appreciation of the benefit arising from systems that aid reader detection and more recently this has related to the evolution of computer aided detection (CAD) [111-113]. Eyetracking has helped understanding of how CAD marks impact on visual search behaviour[114]. In mammography, eye tracking studies have been utilised to attempt to improve the detection accuracy of a CAD computer algorithm and to demonstrate the use of CAD for identification of specific lesion characteristics missed by the human observer. These have shown the strength of CAD, compared to the human reader, for identification of atypical microcalcification clusters [115].

Prevalence

Gaze characteristics variation with recall / repeat viewing has been demonstrated, even over extended ‘washout’ periods [105]. This is important, as it calls into question the legitimacy of using repeated viewing of images under differing conditions as methodological approach for imaging studies.

CTC

There have been many developments around 3D presentation of medical imaging in recent years, CTC being one of the most notable. Volumetric data acquisition presented in a novel 3D visualisation presents the observer with search and interpretation tasks that are neither described nor understood and, previous to the work described in this thesis, nor studied with eye tracking technology. The feasibility of using eye tracking to better understand gaze when performing 3D CTC had been initiated at our unit. Phillips et al

had demonstrated that gaze data could be obtained from fly-through 3D CTC videos by using a circular mask around each polyp to identify its location on each individual video frame. Subsequently relating gaze location to the edge of the mask (as a surrogate of polyp location) allows gaze tracking of a moving 3D image. Observers used a single, non-targeted mouse click to record whether they believed a polyp was present and when it appeared [109]. This method of both experimental design and analysis worked well and forms the basis of the downstream studies described in this thesis. The author of this thesis hoped that eye tracking would present an opportunity to better understand the complex task of CTC interpretation and enable a framework to be developed for the use of gaze tracking moving 3D images in future research studies.

Section B: Understanding reader behaviour and performance in CTC through the utilisation of eye-tracking

Chapter 3: Developing metrics that describe eye-tracking of 3D moving images.

Overview

- Our group has shown eye-tracking during the interpretation of CTC images to be technically feasible but no metrics have been developed that describe the visual perception task in this new paradigm or how varying study and reader factors may impact on diagnostic accuracy.
- This section employs eye-tracking in CTC to tackle a range of reader and study variables that have been hypothesised or proven to affect performance in CTC or more generically, in imaging studies.
- It builds on and develops new eye tracking techniques from our research team that allow us to examine where readers' gaze falls when reading a 3D medical image, where the target pathology is both moving and changing in size.
- To allow robust and consistent analysis of eye-tracking techniques in these studies, a framework is first described by which data can be categorised and analysed.

This research has been published as:

Towards a framework for analysis of eye-tracking studies in the three dimensional environment: a study of visual search by experienced readers of endoluminal CTC.

Helbren E, Halligan S, Phillips P, Boone D, Fanshawe TR, Taylor SA, Manning D, Gale A, Altman DG, Mallett S. Br J Radiol. 2014 May;87(1037):20130614. doi: 10.1259/bjr.20130614. Epub 2014 Feb 20. PMID: 24689842

Introduction

Metrics such as “time to first hit” and “dwell time” are well-established measurements used to compare the performance of different observers in 2D environments [102, 116-118]. However, multi-planar imaging, 3D reconstructions and endoluminal “fly-through” viewing all demand patterns of visual search that are more complex than those associated with the interpretation of 2D displays. In particular, the 3D image is often moving and so gaze strategies are more akin to looking at a “video” than at a static image.

Methods to obtain gaze-tracking information from readers of moving 3D medical images have been described recently [109]. However, standard metrics for analysis of visual search that have been derived from static 2D images might not be applicable to new 3D display paradigms, especially where pathology is often both moving and changing in size during display, situations that are not encountered in conventional 2D gaze-tracking. Using experienced readers to observe videos obtained from 3D endoluminal CTC

examinations, we aimed to develop a range of measurements intended specifically to permit investigation of visual search, recognition, and decision making in the 3D environment. Our intention was to propose a comprehensive framework of metrics suitable for application in the 3D paradigm that builds on previous initial work [109].

Methods and Materials

CTC cases collected during two prior studies [1, 2] were used to obtain eye-tracking data from volunteer readers. All readers gave informed written consent and were assigned a reader number at random. The reader number was then used as the sole identifier for each reader, allowing the eye tracking data collected to be recorded and held anonymously.

Video preparation

Eight endoluminal fly-though videos, each of 20s duration, were recorded from 3D CTC fly-through examinations viewed on a Viatronix® V3D colon imaging workstation (Viatronix Inc., Stony Brook, NY). Patient cases were selected from a bank of CTC studies used for previous research related to interpretation of CTC by both experienced and novice readers.[1, 2] Studies included those from symptomatic and asymptomatic patients accrued from four US and three European centres. All studies had both prone and supine acquisitions following full bowel purgation and colonic insufflation. A reference truth as to the location and size of polyps on each patient case had been established previously by three radiologists experienced in CTC interpretation in consensus (Professor Steve Halligan, Professor Stuart

Taylor and Dr. David Burling) with the aid of the original radiological, endoscopic and pathological reports.

Cases were selected by a radiologist (Dr Darren Boone) with experience of >500 endoscopically validated cases, to obtain a subset of videos that were neither “too easy” nor “too difficult” to interpret. Consulting data from the previous reader studies, 20 cases were obtained in which a false-negative or -positive diagnosis of a polyp had been made previously by approximately 50% of experienced readers. Cases were then excluded if the target polyp could not be demonstrated on either endoluminal projection or if it was within 5 seconds navigation of the rectal ampulla or caecal pole during endoluminal fly through. Where the polyp was visible in both prone and supine acquisitions, the least conspicuous acquisition was selected. Five videos with true-positive polyps (with diameters of 6, 8, 11, 12 and 25mm respectively) were selected. Three videos with prior false-positive polyps (with estimated diameters of 5, 7 and 10mm respectively) were also selected to provide true- and false positive lesions in a 2:1 ratio.

Readers

Eye-tracking data was collected from ten experienced readers who were the teaching faculty at a CTC “hands-on” workshop (ESGAR Amsterdam workshop, April 2010). “Experienced readers” were defined as radiologists who had previously interpreted >300 CTC studies independently. All readers were unaware of the prevalence of abnormality within the dataset prior to viewing, and no feedback was given regarding their diagnostic performance.

Eye tracking

An infrared eye tracker (Tobii X50®; Tobii Technology, Danderyd, Sweden) was positioned beneath the viewing screen, and Studio™ capture software (Tobii Technology) was hosted on a laptop. Eye-tracking accuracy was 0.5° and 20 screen pixels at approximately 60cm viewing distance. The video area was 512 x 512 pixels. Eye tracking was overseen by an image perception scientist (Dr Peter Phillips) with 8 years of experience.

Videos were viewed during the workshop, in a quiet area of a reporting room that was designated specifically for this study. Following a five-point calibration exercise, a “warm-up” video was used to assess the ability of the eye tracker to obtain sufficient data and to familiarise readers with the procedure. Readers wore glasses/contact lenses as per normal. Each reader held a computer mouse prior to commencing each video. The following instructions were then displayed onscreen: “You are about to be shown some short video clips of fly-throughs. Some will have pathology and some will not. Please click the mouse when you see a lesion which you consider highly likely to represent a real polyp or cancer”. Readers were told to click once for each lesion they identified.

The eight videos were then presented to readers in a randomised order. Readers took approximately 10 minutes to complete the setup described above and to view all videos. Eye-tracking data and number/ timing of mouse clicks were recorded for all ten readers viewing all eight videos. Readers were not required to target any polyp with the mouse; they simply clicked to indicate their belief that pathology might be present on the 3D fly through.

Data preparation and display

Following data collection, a circular region of interest (ROI) was applied around the polyp on each individual video frame where the polyp (true- and false-positive) was visible. This was then related to the readers' gaze for each individual frame by calculating the distance from the gaze point to the closest ROI boundary point for each point of gaze data acquired during the time the polyp was onscreen (Figure 4). Gaze points recorded within 50 pixels beyond the outer rim of the ROI were considered to have fixed upon the polyp to ensure that all gaze directed at the polyp was captured. This represented a 1.25° visual acceptance radius,[109] where the ROI boundary fell within very high visual acuity [119]. For each reader, the distance from the gaze point to the ROI boundary was plotted against time, for the duration that any polyp was onscreen, and included the identification time (mouse click), if any. A representative graph is displayed in Figure 5.

Figure 4: Eye tracking data preparation

a) Colonic polyp annotated with a black circular region of interest (ROI). Gaze point demonstrated as blue circular dots. b) Diagrammatic representation with blue lines showing distance of gaze point to ROI.

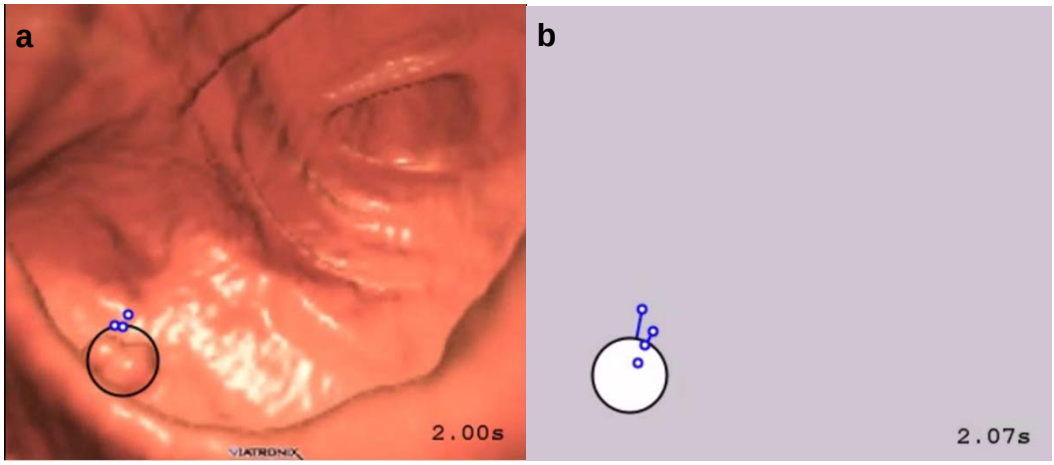
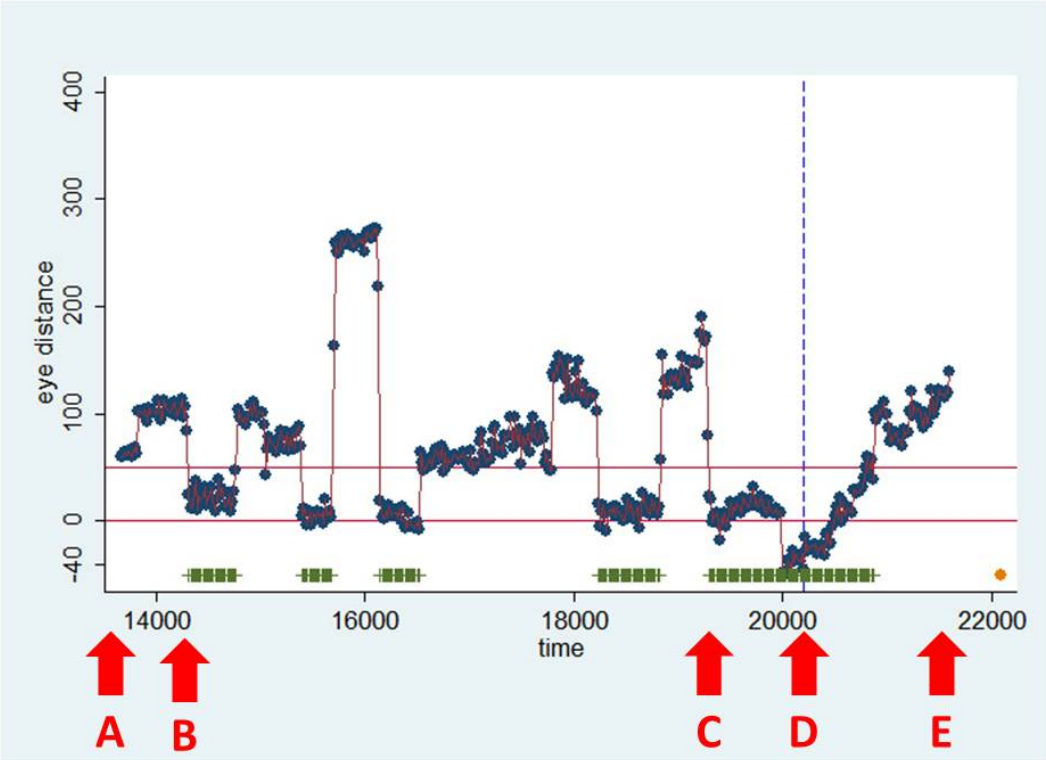


Image courtesy of Dr P Philips

Figure 5: Gaze metrics

The graph shows the distance in pixels (y axis) between individual gaze points (dots) and the region of interest (ROI) drawn around the polyp after viewing, plotted against time (milliseconds, x axis). The horizontal lines indicate the 50 pixel margin to the ROI boundary. Thus, a gaze point falling within this boundary denoted the observer looking at the polyp. Consecutive data points lasting > 100 milliseconds within this margin were defined as gaze pursuits (highlighted by horizontal green bars). The vertical dashed line represents the timing of a mouse click (observer identification of the polyp). The following labels denote key events: (A) point at which the ROI first becomes visible onscreen; (B) time at which pursuit is first recorded; (C) time at which the final pursuit immediately preceding polyp identification begins; (D) time of mouse click; and (E) time at which the ROI leaves the screen.



Statistical analysis

Eye pursuits were defined when within 50 pixels from the polyp ROI boundary and lasting for at least 100 milliseconds. Missing data were imputed using multiple imputation methods [120] adapted for longitudinal

data. A pursuit distance >50 pixels was characterised as a pursuit termination, provided that (to allow for measurement error) either the average pursuit distance of four to six near-contemporaneous gaze points was >50 pixels or the observed pursuit distance was more than two standard deviations of measurement error greater than the average pursuit distance within that pursuit. Gaze metrics were defined as in Figure 5 (time to first pursuit corresponding to A to B; overall assessment time A to E); pursuit time being total time within a 50 pixel distance from the ROI boundary. For each metric, data sets with either >50% missing data or at least 1 block of 50 consecutive missing observations were examined to identify the values that would be unreliable and should be excluded from the analysis.

For metrics that measured time to an event, e.g. time to first pursuit, the event was censored if the event did not occur and was truncated at the time point when the event was no longer possible. For eye pursuits, the censor time was defined as occurring when the ROI was no longer visible (i.e. when the polyp left the screen), and at 500 milliseconds after this for events involving time of ROI identification. Data were analysed using STATA® 12 (StataCorp, College Station, TX). Median and interquartile ranges (IQRs) were calculated to summarize percentage assessment pursuit time across readers and cases.

Results

For clarity, metrics are defined by referring to one reader's eye tracking of a visible ROI in a single video (Figure 5). Additionally, Table 3 details two readers viewing the same video with corresponding gaze graphs in Figure 6;

this example demonstrates how search varies between readers and how the metrics reflect this.

Pursuit: We defined “pursuit” as a consecutive contiguous gaze point related to the ROI (i.e. within a boundary distance of 50 pixels) and lasting for 100 milliseconds or longer.

Time to first pursuit: Since “time to first fixation” is a commonly used outcome for 2D gaze-tracking, we defined “time to first pursuit” as the time elapsing between the first onscreen appearance of the ROI and commencement of first pursuit, if any (Figure 5). Both readers pursued the polyp soon after it became visible (at 0.42 and 0.57seconds), both within 10% of the total onscreen time.

Identification and assessment time: To distinguish different components of identification and decision time, we defined the following three features and extracted them from the gaze data, in seconds and milliseconds (Table 3, Figure 5):

- ***Identification time span:*** time elapsing between first onscreen appearance of the ROI and the time of polyp identification (if any) by the reader (represented by the mouse click); time A to D on Figure 5.

- ***Total assessment time span:*** time from first fixation on the ROI (if any) to identification by the reader; time B to D on Figure 5.

- ***Last assessment time span:*** time from commencement of the ROI fixation immediately preceding identification to the time of identification; time C to D on Figure 5.

A reaction time of 500 milliseconds was included to capture mouse clicks occurring very soon after the polyp left the screen.

Pursuit time: We identified two different components of pursuit time. We expressed both as a percentage of the total onscreen time during the periods described:

- ***Assessment pursuit time:*** the aggregated time for individual pursuits (if any) occurring prior to polyp identification; the summed length of the green horizontal bars on Figure 5 prior to the mouse click.

- ***Total pursuit time:*** the aggregated time for individual pursuits (if any) occurring for the total onscreen time of the ROI (i.e. including pursuits occurring after identification); the summed length of the green horizontal bars on Figure 5.

Region of interest size: We expressed the size of the ROI as a percentage of visible video area at crucial points during reader gaze as follows:

- ***size at first pursuit****:* B on Figure 5

- ***size at longest pursuit****:* C on Figure 5.

Readers pursued polyps at relatively small sizes, as a percentage screen area; 0.26% and 0.31% at the first pursuit and 0.69% and 2.38% at the longest pursuit. In our example video, the largest polyp size is 8.54%, indicating that both the first and longest pursuits were at relatively small polyp sizes. The polyp size when readers clicked was much larger, at 2.90% and 7.10%.

Pursuit frequency: We identified two different components of pursuit frequency expressed as the rate of pursuits per second. This facilitated comparison across videos where onscreen ROI time will vary.

- ***Assessment pursuit rate:*** the rate of individual pursuits (if any) occurring prior to the point of polyp identification; the number of individual horizontal bars per second on Figure 5 prior to the mouse click.

- ***Total pursuit rate:*** the rate of individual pursuits (if any) occurring for the total onscreen time of the ROI (i.e. including pursuits occurring after identification); the number of individual horizontal bars per second on Figure 5.

Analysis of metrics across readers: The total time the ROI was visible varied from 2.47 to 8.87 seconds, with a median of 5.73 seconds (IQR, 3.05 to 7.93 seconds). The length of time which the polyp was onscreen varied between videos. To allow the comparison of videos, metrics affected by this variation where expressed as a percentage of total on screen time (see table 3). To illustrate the power of these metrics to summarise visual search patterns across readers and cases, we present results that assess pursuit time across all ten readers and eight videos: the median assessment pursuit time was 43% (IQR, 23 to 53%).

Figure 6: Differences in gaze example

Gaze graphs for two readers demonstrating the differing visual search characteristics that can be seen when viewing the same video. Metric values for the same readers are summarised in Table 3.

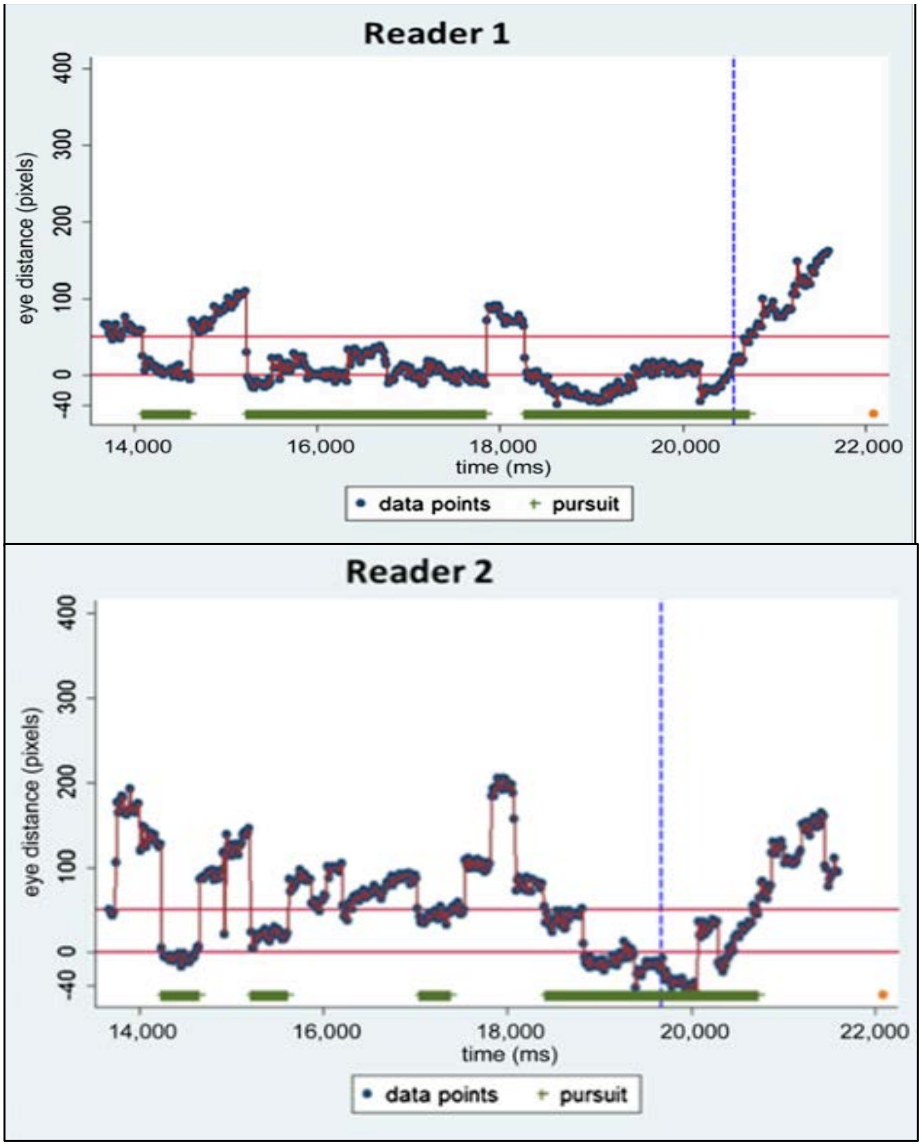

Reader 1
eye distance (pixels)
-40
0
100
200
300
400
14,000
16,000
18,000
20,000
22,000
time (ms)
data points
pursuit
Reader 2
eye distance (pixels)
-40
0
100
200
300
400
14,000
16,000
18,000
20,000
22,000
time (ms)
data points
pursuit

Table 3: Metrics applicable to eye tracking of three dimensional moving studies

Metric values are presented for two readers reading the same video, corresponding to the gaze graphs in Figure 6.

Metric category	Metric descriptor	Reader 1 (seconds) (% of total time ROI onscreen)	Reader 2 (seconds) (% of total time ROI onscreen)
Time to first pursuit	Time to first pursuit	0.42 (5.2%)	0.57 (7.1%)
Identification and assessment time	Identification time span	6.87 (87%)	5.99 (76%)
	Total assessment time span	6.46 (81%)	5.43 (68%)
	Last assessment time span	2.28 (29%)	1.26 (16%)
Pursuit time	Assessment pursuit time	69%	30%
	Total pursuit time	71%	43%
ROI size (percentage of visible video area)	Size at first pursuit	0.26%	0.31%
	Size at longest pursuit	0.69%	2.38%
Number of pursuits	Assessment pursuit rate	$0.44s^{-1}$	0.67^{-1}
	Total pursuit rate	0.38^{-1}	0.5^{-1}

Discussion

When viewing an image, features of interest are brought into the centre of field of view via “foveal fixation”, providing the sharpest visual detail in the region of conscious attention. Multiple fixations (“spatial clustering”) imply a feature of particular interest [90]. It is relatively straightforward to record the location and duration of foveal fixations for 2D medical images.[92, 117] By contrast, when images are moving, we fix and follow objects using rotational eye movements to stabilize the fovea on the target. Because both the image and the location of any fixed feature change frame by frame, and the nature of eye movements involved is different, the simple x, y co-ordinates and “heat maps” (for fixation duration) used to represent visual search in 2D images can no longer be applied. Using CTC as an example of the 3D medical imaging paradigm, we sought to develop a comprehensive range of metrics intended to facilitate investigation of visual search, recognition, and decision making in the 3D environment, building on descriptions by our group priot to this thesis.[121] By relating gaze to an ROI in terms of proximity and time, we derived a set of metrics applicable to a wide range of 3D imaging paradigms, basing these on parameters already established for 2D studies. We considered the unique qualities of the 3D environment, e.g. feature variation between individual frames (expressed via metrics describing the ROI size), and the time-dependent nature of the viewing task. Time pressure is irrelevant for static images, and the location and nature of background features are also constant. Care was taken so that our metrics were potentially applicable to readers of all experience and would extend to studies using different software and eye-tracking systems.

In 2D gaze tracking, a “hit” occurs when readers gaze at lesions directly for a specified minimum time period. Spatial clustering of fixation points over a static ROI is known as “dwell time” and their individual summation as “cumulative dwell”. In 3D gaze tracking, readers’ eyes must follow the ROI across the screen. Assessment is then reflected by time spent “pursuing” the ROI, which we propose as the 3D surrogate of 2D dwell time. We defined pursuit as occurring when readers’ uninterrupted gaze was within a moving ROI boundary for 100 milliseconds or more. In 2D, the number of fixation clusters associated with an ROI has been shown to correlate with identification of true-positive lesions.[116] We were able to identify the number of individual pursuits in 3D and measure their individual and summated duration, with the expectation that this could examine any relationship between repeat pursuits and lesion identification in future studies. It is possible that the time-limited nature of lesion identification in 3D will enhance the importance of such metrics.

Many 2D eye-tracking studies allow readers to control the total time an image is displayed and viewed. In our study, readers could not control display time, since polyps appeared on the screen and disappeared subsequently at pre-determined points. In fly-through 3D, the observer does not change case once a lesion has been detected nor can he/she eliminate a lesion from view once it has been characterized. It is therefore desirable to separate pursuit frequency and times into those that occur before and after lesion identification. We achieved this using a mouse click. Pursuits prior to any click are probably related to lesion recognition and decision. Not all viewers will identify an abnormality as such, so we believe a metric that

reflects total pursuit time when they occur, while the ROI is on-screen, is important. Alternatives to a click, such as verbal response, are possible but we did not investigate this and may be less accurate in terms of timing measurements.

Time to identify an onscreen lesion was considered to reflect both “viewing time”, as for 2D studies, as well as providing insight regarding visual search. Unlike 2D, where viewing time essentially terminates interrogation, noting “decision time” in 3D allows the observer to indicate that a lesion has been identified and to then continue searching for further abnormalities present in the remaining video. If no abnormality is apparent elsewhere, the observer may return to pursue an abnormality already identified. The mouse click allowed us to separate pursuits relating to search and detection. A marker of lesion identification distinct from pursuit was also necessary, because it is possible that experienced readers may perceive abnormalities via peripheral vision, without formal pursuit. This is particularly important where moving images are concerned.

Whereas lesions remain unchanged in size and location in 2D, 3D necessitates a complex ROI that changes in size and position frame by frame. We therefore hypothesised that the ROI size (in pixels) at time of first pursuit, and immediately prior to identification, will aid understanding of 3D perception. Experienced readers might pursue and identify smaller ROIs than would novices or, alternatively, they might appreciate that resolution is maximal when a potential abnormality reaches the image foreground (i.e. just prior to it leaving the screen). This might delay identification time to extract maximal visual information. We were mindful that our methods should be

transferable; an ROI can be created for different and complex lesion morphologies.

Our study has limitations. We investigated endoluminal fly through in an automatic mode. In clinical practice, readers can adjust navigation speed and also stop to inspect potential abnormalities manoeuvers that were not possible within our implementation. A circular ROI was convenient; we adjusted the diameter to represent change in the polyp screen size over time of approach. However, irregular polyps and those seen in profile are more difficult to characterise in this way. Boundary accuracy could be improved via more representative descriptions but this will increase complexity. The 50 pixel threshold was constant for all polyps, which may have entailed distant polyps being defined as “seen” too early. Possible perceptual errors would then be classified as recognition errors. A threshold based on a fixed proportion of the ROI would have the opposite effect. Future thresholds should account for all polyp sizes. We investigated only experienced readers, since our aim was simply to derive a set of metrics applicable to 3D gaze tracking; ultimately, we hypothesise that these metrics will facilitate comparisons between experienced and inexperienced readers (see Chapter 4; pages 72 to 93) that might reveal factors associated with correct lesion detection that can be used to inform training schedules. We used both true- and false-positive polyps. False-positive polyps were visible abnormalities (e.g. residues) that had been consistently labelled as polyps by experienced readers in previous research studies. In the future, we intend to determine if there are any gaze characteristics that differentiate these from true-positive polyps. While we believe this work represents important steps towards 3D

gaze tracking, further work should investigate gaze when the ROI is off screen and for multiple simultaneous ROIs. We will also use these metrics to examine how observers' gaze is affected by the presence of computer assisted detection marks on the screen (see Chapter 5; pages 94 to 114). We present only one metric summarised across all readers and cases. Future work will use multilevel analyses accounting for clustering of data within readers and cases, hence allowing use of time to event survival analysis and count data.

In summary, this work presents a comprehensive range of metrics applicable specifically to studies of eye tracking in the 3D paradigm, including where potential lesions are both moving and changing in size. The author believes these metrics provide a reproducible framework to investigate 3D visual search and are potentially applicable to a wide range of research studies performed in this new, exciting environment. These metrics should facilitate identification of factors related to expertise in interpretation of 3D medical images.

Chapter 4: Investigating influences on reader search and performance in 3D CTC: Experienced versus inexperienced readers.

Overview

- Visual perception patterns across groups of inexperienced and experienced readers were captured while viewing 3D CTC.
- Inexperienced readers identified a lower percentage of lesions than experts (67% vs 75% respectively) but exhibited a similar percentage of eye pursuits on lesions (97% vs 96% respectively). This suggests that error arose in either the decision making phase of the visual assessment.
- Time to first pursuit was significantly shorter for experienced readers (hazard ratio, 1.22 [95% confidence interval: 1.04 to 1.44]; p= 0.017). Other metrics were not significantly different.
- During the time polyps were on screen, readers spent approximately one-third of the time pursuing polyps and two-thirds looking at other areas of the video.

This research has been published as:

Tracking eye gaze during interpretation of endoluminal three-dimensional CTC: visual perception of experienced and inexperienced readers.

Mallett S, Phillips P, Fanshawe TR, **Helbren E**, Boone D, Gale A, Taylor SA, Manning D, Altman DG, Halligan S. Radiology. 2014 Dec;273(3):783-92. doi: 10.1148/radiol.14132896. Epub 2014 Jul 15. PMID: 25028782

Introduction

Experienced readers of CTC are more diagnostically accurate than less experienced readers [1, 2, 63], but our knowledge of why this is, is limited. Better understanding the errors made by inexperienced readers should allow us to improve training for them.

When viewing 2D imaging a ‘global and focal’ model has previously been described in which there is an initial fast global impression taken by the reader that then allows more focal searches of regions of interest to be undertaken. Experienced readers characteristically have an emphasis on global recognition rather than focal search, with experts appearing to be more effective in making correct decisions at an earlier stage [95, 122, 123].

It is thought that most false-negative errors, when viewing 2D formats, are due to errors of recognition or decision, rather than an inability to scan the image appropriately [92, 93, 124]. The nature, frequency or importance of false-negative interpretation error when viewing 3D studies has not been described. Chapter 3 developed and described a range of metrics that measure visual perception in the 3D paradigm so that gaze behaviour can be compared across readers in multi-reader, multi-case studies. By examining visual search patterns of experienced and inexperienced readers of CTC, we aim to identify key stages of image perception, and examine how these stages relate to false-negative errors when viewing a moving 3D image.

Materials and Methods

Video data and readers

Twenty-three 30 second video clips were recorded by the author of this thesis or Dr Darren Boone from automated 3D CTC fly-through examinations viewed on a medical imaging workstation (Vitrea; Vital Images, Minnetonka, Minn). These cases were taken from the same bank of anonymised CTC studies as used in Chapter 3 and referenced in the ethics section (page 15). Two sets of videos were used, with each video including one polyp or polyp-like feature. Polyps that were determined to be false-positive findings but were indistinguishable from polyps that were true-positive findings, and so were treated as polyps. Reference standard decisions had been made by a consensus panel based on majority decision of three experts by using radiologic, endoscopic, and pathologic reports [1, 2]:

Set a

8 videos, of polyps from 5 to 25 mm in diameter, recorded by Dr Darren Boone were used, which had been used in a previous study [109]

Set b

15 videos (of polyps 5 to 7 mm in diameter) were recorded by the author of this thesis. For each video, a start point was chosen that positioned a single polyp randomly, between 5 and 25 seconds after the video start point. Cases used in previous studies where neither intentionally included nor excluded. Where cases were used that had been included in the prior study (where shorter video clips were used) the clips were re-recorded. Re-recording clips

generated different orientations of the fly-through video meaning that there was no direct repetition of images used.

Sixty-five radiologists participated. Twenty-seven experienced readers (defined as >200 prior cases) were recruited at a subspecialty conference (2012 European Society of Gastrointestinal and Abdominal Radiology Annual Meeting) or because they were faculty at a CTC “hands- on” workshop (2010 European Society of Gastrointestinal and Abdominal Radiology Annual Meeting, Amsterdam). Thirty-eight inexperienced readers (defined as <199 prior cases) were recruited at a sub-specialty conference (2012 European Society of Gastrointestinal and Abdominal Radiology Annual Meeting) or at a UK teaching hospital (University College Hospital, London, England).

Eye gaze tracking

Eye tracking was overseen by an image perception scientist (Dr Peter Phillips). Videos were viewed consecutively in different random order by each reader, in sets of seven or eight videos with an optional few minutes rest between. Reading occurred in a quiet room with constant, ambient light. The video clips were shown on a liquid-crystal display monitor (SyncMaster 723N; Samsung, Suwon, Korea; 1280 X 1024 resolution, 1 pixel = 0.264 mm) approximately 60 cm in front of the reader. Videos measured 384 X 384 pixels (10.1 X 10.1 cm), with a visual angle of 9.6° [109]. We placed an eye tracker (Tobii X50 or Tobii X120; Tobii Technology, Danderyd, Sweden) beneath the viewing screens, and eye movements of readers were recorded without use of a head rest.

Data collection began with an initial five point calibration and warm-up video as previously described in Chapter 3. As in Chapter 3, readers again used a computer mouse click to indicate when the they saw a lesion that they considered to be highly likely to represent a polyp or cancer.

No information was given to the readers regarding the number of videos with polyps or number of polyps per video.

Data preparation and eye metrics

After eye gaze tracking, a circular ROI was drawn around each polyp on each video frame by Dr Peter Phillips, as shown in Figure 4, page 59.

Outcome metrics were as previously defined in Chapter 3 and Figure 5 (page 60 to 64) with the only adaptation being to the ROI size at longest pursuit metric. In this study, percentage size of ROI at the start of the longest pursuit was only calculated when readers indicated a lesion was present, taking the size of the ROI at the start of the longest pursuit before the first click. This differs from the ROI size at longest pursuit calculated in Chapter 3, which was for the longest pursuit during the time the ROI was onscreen, regardless of whether an identification occurred. The intention was to try and identify differences in the ‘recognition’ phase of visual search (as explained in Chapter 2, page 51) between readers of differing experiences.

Statistical analysis

Statistical analysis was performed by statisticians Dr Sue Mallett and Dr Tom Franshawe. Missing data was imputed by using multiple imputation methods [120] adapted for longitudinal data. For each metric, datasets with either

more than 50% missing data or at least one block of 50 consecutive missing observations were examined to exclude unreliable observations. Subject specific multi-level modelling with cross-classified analysis used random intercepts for reader and case and used fixed effects for reader experience (Stata12; StataCorp, College Station, Tex). These models allowed data interpretation in terms of individual readers, rather than by averaging across the population of readers. Reader groups were compared within the same videos, which accounted for video-specific effects. Proportional hazards models were applied for survival variables, time to first event measures where data-set included values censored by computer mouse click (time-to-first pursuit, total assessment timespan, last assessment period by using software [xtmepoisson, Stata12; StataCorp]), linear models for continuous variables (xtmixed, Stata12; StataCorp), logistic models for binary variables (xtmelog- it, Stata12; StataCorp), and Poisson models for count variables (xtmepoisson, Stata12; StataCorp). Time-to-first- pursuit analysis used a shared frailty Cox model with single level for case because models that included random effects for readers and models fitted with xtmepoisson did not converge.

When the decision to identify a potential polyp or cancer was important to understand visual perception behaviour, a binary fixed-effect variable was used to adjust for correct mouse clicks to allow reporting of results separately. Presentation of typical values used median and interquartile ranges across readers and cases for survival variables and when data were not normally distributed. To estimate how metrics changed according to the total time a polyp was on the screen, the length of time that the polyp was

visible was included in models as a fixed effect. We pre-specified three of the primary outcome metrics developed in the prior Chapter 3 (pages 62 to 64) to compare reader groups: time to first pursuit, assessment pursuit time, and assessment pursuit rate. Confidence intervals (CIs) were calculated for means and interquartile ranges (IQRs) were calculated for medians.

Results

Of the 65 radiologists, 23 radiologists interpreted eight videos (10 radiologists were experienced, 13 were inexperienced) and 42 interpreted 15 videos (17 radiologists were experienced, 25 were inexperienced). In total, 27 readers were experienced in CTC and 38 were inexperienced (Table 4). We recorded 803 observations where polyp identification was made, of which 787 had successful tracking of eye gaze. Many observations had small proportions of longitudinal data missing for eye position when a polyp was visible (causes included blinking, looking away from the screen, stray reflections and head movement); multiple imputation methods to handle missing data allowed for calculation of metrics for almost all observations. For each individual metric, a small number of observations were though excluded because of the timing of missing data that it was felt meant the use of the data unreliable.

Experienced readers were significantly more likely to identify polyps than were inexperienced readers (odds ratio, 2.11 [95% CI: 1.03, 4.34]; p = 0.042). Across all readers and cases, experienced readers had higher rates of polyp identification than inexperienced readers (75% [250 of 334] vs 67% [314 of 469], respectively).

Table 4: Demographic information for experienced and inexperienced readers

Parameter	No. of Inexperienced Readers (n=38)	No. of Experienced Readers (n=27)
Men	22 (58)	22 (81)
Wore glasses	11 (29)	8 (30)
Wore contact lenses	4 (11)	7 (26)
Consultant	14 (37)	21 (78)
Specialist registrar	22 (58)	6 (22)
Other	2 (5)	0 (0)
Gastrointestinal specialty	5 (13)	23 (85)
No. of cases seen previously		
0 to 9	30 (79)	0 (0)
11 to 49	6 (16)	0 (0)
50 to 199	2 (5)	0 (0)
200 to 299	0 (0)	2 (7)
>300	0 (0)	25 (93)

Note - Date in parentheses are percentages

Identification and assessment time

97% (760 of 787) of videos elicited a polyp gaze pursuit (Table 5). The percentage of pursuits was similar for both experienced and inexperienced readers. On average, polyps were on screen for 7.14 seconds (range, 2.38 to 18.99 seconds).

Table 5: Polyp pursuit and identification by experience

Reader experience	Gaze pursuit of polyp	Click indicative of polyp seen
Inexperienced (%)	97 (455/468)	67 (314/469)
Experienced (%)	96 (305/319)	75 (250/334)
Total (%)	97 (760/787)	70 (564/803)

Note - Data in parentheses are numerator and denominator.

Time-to-first-pursuit was significantly lower for experienced readers, indicated by a higher hazard ratio (hazard ratio, 1.22 [95% CI: 1.04 to 1.44]; p = 0.017), with a median of 0.62 seconds (IQR, 0.38 to 1.06 seconds) for experienced readers and 0.66 seconds (IQR, 0.37 to 1.32 seconds) for inexperienced readers. When there was no polyp identification, the first fixation of the polyp occurred earlier.

Time-to-event analysis was reported for the identification and assessment times with high hazard ratios indicating a shorter span of time. There was no significant difference in the identification time span for a polyp between experienced and inexperienced readers (hazard ratio, 1.17 [95% CI: 0.98,

1.39]; p = 0.073) (Table 6). Experienced readers had a median identification time span of 5.23 seconds (IQR, 2.36–11.43 seconds) and inexperienced readers had a median identification time span of 5.64 seconds (IQR, 2.52–18.35 seconds).

Table 6: Summary of metrics for experienced and inexperienced readers

Parameter	Experienced readers	Inexperienced readers	Comparison of metrics (CI)	p Value
Median time to first pursuit (sec)	0.62 (*IQR* 0.38 to 1.06)	0.66 (*IQR* 0.37 to 1.32)	*HR* 1.22 (1.04, 1.44)	0.017
Median identification time span (sec)	5.23 (*IQR* 2.36 to 11.43)	5.64 (*IQR* 2.52 to 18.35)	*HR* 1.17 (0.98, 1.39)	0.073
Median total assessment time span (sec)	3.95 (*IQR* 1.58 to 7.09)	3.84 (*IQR* 1.63 to 9.45)	*HR* 1.16 (0.89 to 1.51)	0.279
Median last assessment timespan (%)	1.35 (*IQR* 0.72 to 2.84)	1.45 (*IQR* 0.83 to 3.61)	*HR* 1.13 (0.95, 1.35)	0.165
Mean assessment pursuit time (sec)	2.41 (*CI* 1.85, 2.96)	2.16 (*CI* 1.61, 2.71)	0.25 (-0.05,0.52)	0.099
Mean assessment pursuit time (%)	34 (*CI* 29, 39)	31 (*CI* 26, 36)	3 (-1,6,)	0.180
Mean total pursuit time (sec)	3.03 (*CI* 2.33, 3.73)	2.93 (*CI* 2.23, 3.62)	0.11 (-0.16,0.37)	0.436
Mean total pursuit time (%)	43 (*CI* 37, 49)	42 (*CI* 36, 48)	1 (-2,4)	0.550
Mean assessment pursuit rate (per sec)	0.45 (*CI* 0.39, 0.51)	0.45 (*CI* 0.39, 0.51)	-0.0026 (-0.057,0.052)	0.925
Mean total pursuit rate (per sec)	0.55 (*CI* 0.50, 0.60)	0.56 (*CI* 0.50, 0.61)	0.009 (-0.040,0.060)	0.706
Median ROI at first pursuit (%)	0.38 (*IQR* 0.09 to 1.07)	0.48 (*IQR* 0.12 to 1.13)	*OR* 0.91 (0.73,1.13)	0.396
Median ROI at longest pursuit (%)	0.69 (*IQR* 0.34 to 1.55)	0.74 (*IQR* 0.31 to 1.46)	*OR* 0.89 (0.65, 1.25)	0.522

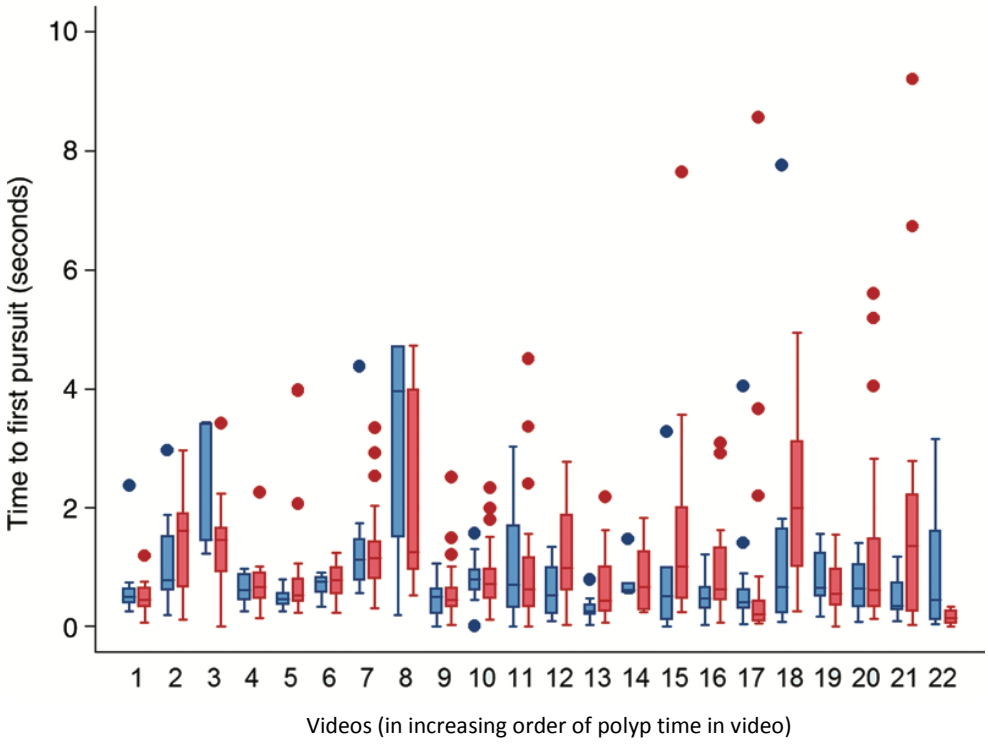
Comparison is by difference in means unless stated as *HR* = Hazard ratio or *OR* = Odds ratio. *IQR* = Interquartile range. *CI* = Confidence interval.

There was no significant difference between experienced and inexperienced readers for total assessment time span (i.e., time from first pursuit to mouse click, hazard ratio, 1.16 [95% CI: 0.89, 1.51]; p = 0.279) (Table 6). Compared with other measures, such as time-to-first-pursuit (Figure 6), total assessment period was significantly influenced by the amount of time that polyps were on screen (Figure 7).

Figure 7: Time to first pursuit with length of time polyp visible on video

No relationship demonstrated.

Experienced readers (blue), Inexperienced readers (red)



The total assessment period was 58% for experienced readers (95% CI: 49%, 66%) and 54% for inexperienced readers (95% CI: 45%, 63%), although there was considerable variation across readers and videos. Individual values ranged from 0.8% to 99.8% across readers, and variation by case was greater still. The total assessment time span was, on average, increased by 0.66 seconds for each additional second the polyp was on screen, Figure 8.

Figure 8: Total assessment period with length of time polyp on video

Total assessment period increases with time that polyp visible in video.

Experienced readers = blue. Inexperienced readers = red.

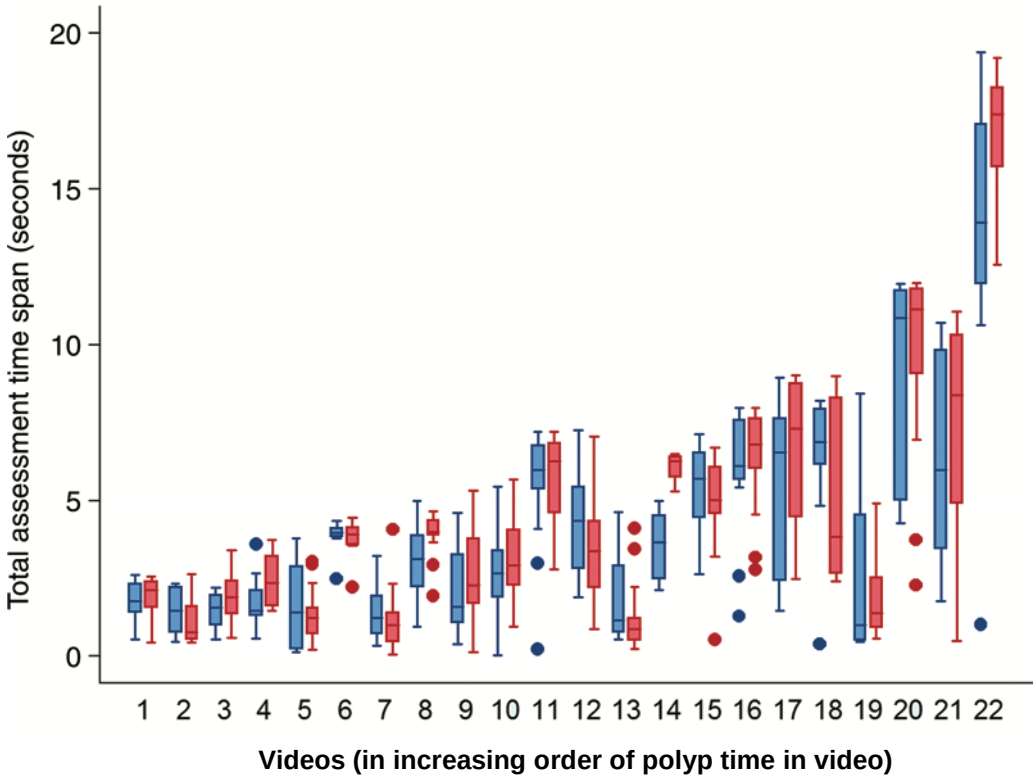


Table 7: Metric change for each second a polyp is visible in video

Metrics significantly changed by time polyp is visible on video. Results are split according to whether readers identified presence of a polyp. Results are averaged over all readers and cases.

Metric	Polyp Identified	Polyp Not Identified
Identification time span (sec)	0.69 (0.64, 0.74)	1.00
Total assessment time span (sec)	0.66 (0.60, 0.71)	1.00
Assessment pursuit time (sec)	0.24 (0.20, 0.28)	0.24 (0.19, 0.29)
Total pursuit time (sec)	0.33 (0.29, 0.36)	0.24 (0.19, 0.29)

Note - Data in parentheses are CI.

There was no significant difference between groups (p = 0.134) for last-assessment time span (i.e., time between commencement of the last individual pursuit to mouse click); median, 1.45 seconds; (IQR, 0.83–3.61 seconds) for inexperienced readers versus 1.35 seconds (IQR, 0.72–2.84 seconds) for experienced readers.

Pursuit time and pursuit frequency

Both groups viewed polyps by using multiple discrete pursuits. Assessment pursuit time (time spent viewing a polyp directly before a decision to click) was reflected as a percentage of the total on-screen time, which allowed for comparison across videos (because polyps were visible for different lengths of time). There was no significant difference between groups, with average assessment pursuit time for experienced readers just 3% more than for inexperienced readers (95% CI: -1, 6; p = 0.180); experienced readers 34% (range, 0 to 88 [95% CI: 29, 39]) versus inexperienced readers 31% (range, 0 to 89 [95% CI: 26, 36]).

Assessment pursuit time and total pursuit time increased significantly with the total time a polyp was visible (Figure 9). For each extra second on-screen, there was an average increase of 0.24 seconds for assessment pursuit time both with and without polyp identification, and, on average, 0.33 seconds for the total pursuit time if a polyp was identified (Table 7). Average assessment pursuit time was 2.16 seconds (95% CI: 1.61, 2.71 seconds) for inexperienced readers and 2.41 seconds (95% CI: 1.85, 2.96 seconds) for experienced readers, whereas the average total pursuit time was 2.93 seconds (95% CI: 2.23, 3.62 seconds) and 3.03 (95% CI 2.33, 3.37) for experienced readers. No significant difference was therefore seen in either metric with p=0.18 and p=0.436 respectively (Table 6).

Figure 9: Visual perception metrics versus time polyp on screen

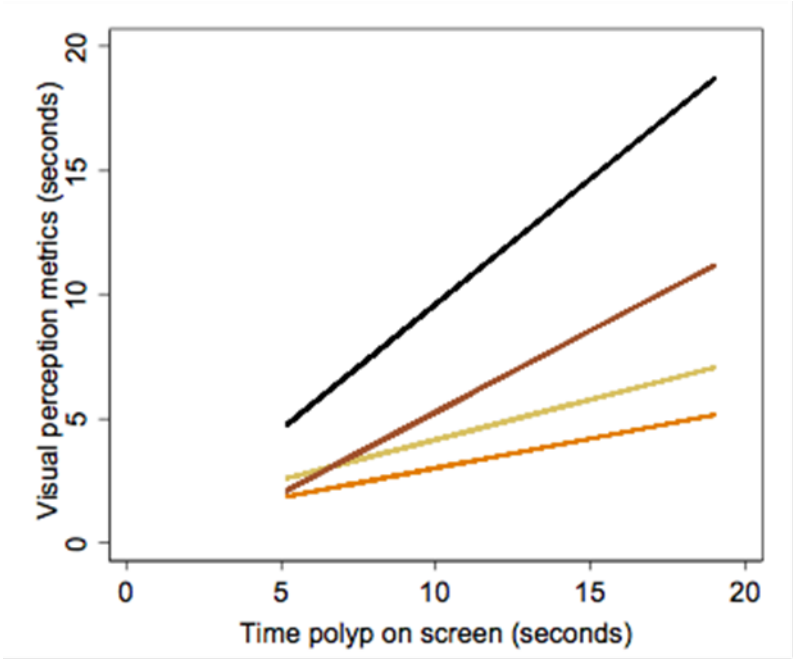
Graph demonstrates the linear increase in assessment and pursuit durations when the polyp was visible for more than 5 seconds

Total pursuit time when no polyp seen (black line)

Total assessment time span (brown line)

Total pursuit time when polyp identified (yellow line)

Pursuit durations (orange line)



We calculated the rate of pursuits per second prior to the polyp identification in the assessment pursuit rate. The average rate was 0.45 pursuits per second (95% CI: 0.39, 0.51) for both experienced and inexperienced readers, with no significant difference. The total pursuit rate (total pursuits of the ROI throughout the video, regardless of identification) was significantly lower when the reader clicked the computer mouse; 0.43 pursuits per second (95% CI: 0.38, 0.48) when there was a mouse click versus 0.49 pursuits per second (95% CI: 0.43, 0.55) when there was no mouse click, giving a difference of -0.06 pursuits per second [95% CI: -0.11, -0.02]; p = 0.008.

ROI size at first and longest pursuits

ROI area at time of first pursuit was analysed to understand what captured readers' attentions first. Data were excluded when there was no pursuit or if pursuit was less than 100 milliseconds from the time of the ROI's first appearance, because this amount of time is insufficient for attention capture and therefore was attributed to chance position of gaze. ROI size at time of first pursuit was not significantly different between groups (Table 3). Average ROI size at time of first pursuit was a median of 0.48% of total video pixel area (IQR, 0.12%–1.13%) for inexperienced and 0.38% (IQR, 0.09%–1.07%) for experienced readers. There was more variation in ROI size at point of attention capture between videos than between readers.

ROI size at the beginning of a reader's longest pursuit before identification of a lesion again showed no significant difference between groups (Table 6) with the median ROI area being 0.74% for inexperienced readers and 0.69% for experienced readers.

Figure 10 demonstrates that no clear relationship is seen between the polyp ROI size and percentage of readers who undertake pursuits. In the graphical example, the percentage of readers pursuing the polyp can be seen to rise prior to the increase in size of the polyp. This lack of relationship between polyp size and percentage of readers pursuing the polyp was seen in almost all cases.

Figure 10: Polyp pursuit and polyp size

Percentage of readers who pursued the polyp (red solid line) and polyp size at each time point (blue dotted line)

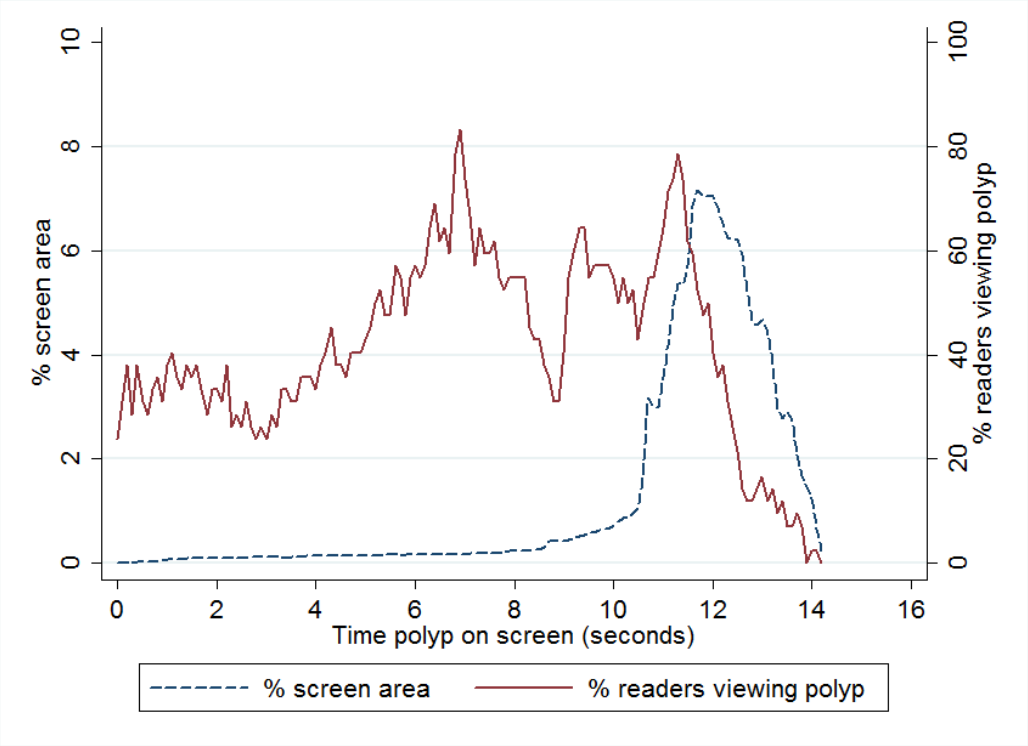
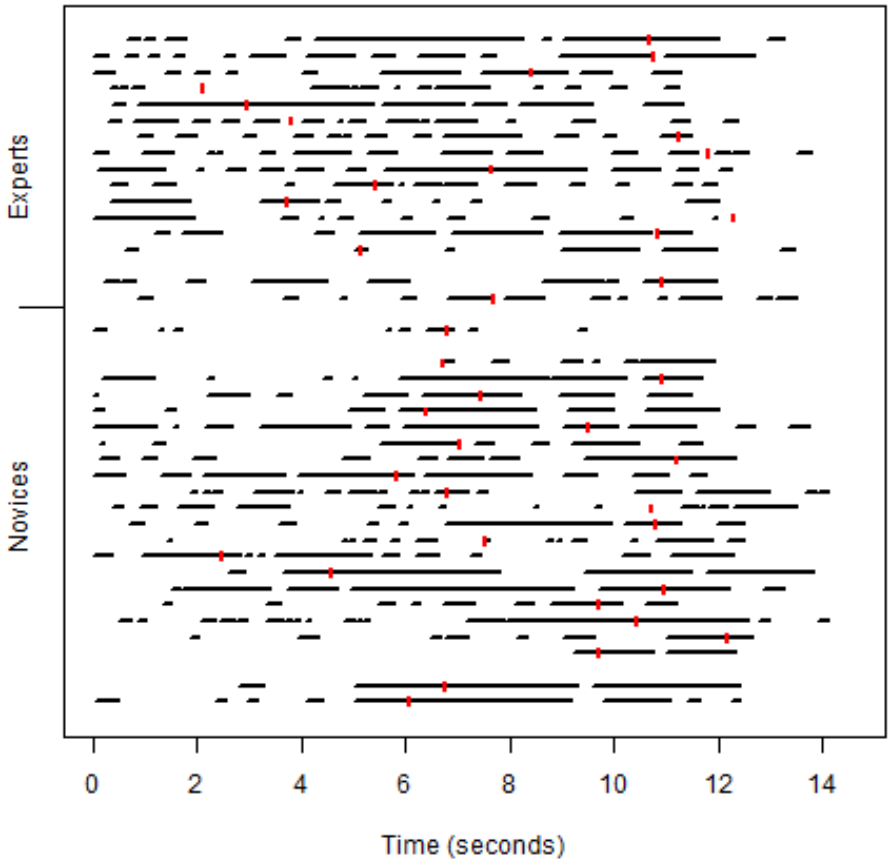


Figure 11 shows the pursuits for individual readers in each row. This shows that at the start of the video most readers used a number of short pursuits, interspersed with viewing other regions of the image. From about 5 seconds, most readers used longer pursuits, often more than 1 second, with short pursuits of other regions of the screen. It can also be observed that readers who pursued the polyp within the first 5 seconds of it becoming visible included those who clicked to identify the polyp at 10 seconds or later. This indicated that delayed identification time was an active process when most readers pursued the polyp.

Figure 11: Individual reader pursuits over time

Eye pursuits for each reader are represented as horizontal black bars. First mouse clicks are represented as a red dash.

Readers grouped as experienced (experts) and inexperienced (novices).



Discussion

Using the metrics developed in Chapter 3 of this thesis we were able to examine visual search behaviour patterns and their variation across readers and videos, comparing experienced and inexperienced groups of readers viewing 3D images. We used 30 second video clips, which were acceptable to readers, and which enabled 787 observations across 65 readers and 23 different videos. Multiple imputation methods were valuable in allowing us to include most recorded observations. Cross-classified multilevel analysis was used to analyse and describe metrics across readers and videos, which allowed for correlations within videos and readers.

Time to first pursuit, a key measure used to distinguish experienced from inexperienced readers in studies of 2D imaging [92, 125] (time to first hit), was the only metric that differed significantly between these groups. However, the absolute time difference was small (experienced vs inexperienced, 0.62 seconds vs 0.66 seconds, respectively), and it is unclear how this influenced polyp identification.

Unlike studies of 2D imaging, we found no significant difference between experienced and inexperienced readers when using analogous metrics suited for 3D moving images; assessment pursuit time, assessment pursuit rate, and identification period (time to decision) [92, 125].

In 2D imaging studies, the commonly used classification for false-negative interpretation is to define the error as occurring in either the scanning, recognition or decision making phases. If we apply this concept to our study then most false-negatives were due to either recognition or decision errors

[92, 93, 124] because 97% of our observations (760 of 787) included a close pursuit of the polyp. Scanning errors can then only account for very few of the polyps missed. The rate of scanning errors was similar for experienced and inexperienced readers, and therefore did not account for the greater accuracy of experienced readers in this study and others [1, 2, 63]. We found that readers almost always examined polyps by a series of multiple pursuits, which suggested that readers recognised a lesion as visually important in most observations, and that the errors lay in the decision to identify the polyp.

Our results suggest that the distribution of errors between scanning, recognition, and decision may vary between different medical imaging tasks. For example, scanning errors have been shown to account for 30%– 50% of errors in visual search for lung nodules in volumetric 2D images of chest CT [114]. It is likely that the distinctiveness of the search object, the complexity of the visual background, navigation, image quality, and speed constraints of the task are all factors in these variations.

We observed that during the identification period, readers spent approximately 40% of the time actively pursuing polyps and 60% looking at other areas of the video. We found small polyp sizes at attention capture at first pursuit (0.48% and 0.38% of screen size for inexperienced and experienced readers, respectively) and at the start of the longest pursuit (0.74% and 0.69%, respectively, for inexperienced and experienced readers). We did not show any relationship between polyp size and number of readers in pursuit of the polyp, with most readers making repeated pursuits including an early pursuit. The last assessment period (i.e. the start

of last pursuit to the first click) may be an important decision time and is on average 1.4 seconds for both groups; interestingly, this did not vary with the length of time polyps were visible on screen, although, the total assessment period was on average 0.66 seconds longer for each additional second a polyp was on the screen.

This study shares many of the limitations already highlighted in Chapter 3 (pages 70 to 71). Briefly, these are; the use of automatic mode endoluminal fly-through where readers are unable to adjust speed or stop and inspect potential polyps as they would in normal daily practice, ongoing concerns regarding the use of a circular ROI. Other boundary descriptions are possible to improve boundary accuracy, but will inevitably be more complex and lastly the use of a set 50 pixel distance threshold constant across all polyp sizes. This can lead to distant polyps as being classified as fixated at a very early stage, when this is possibly not the case. Scanning errors could then be misclassified as recognition errors. An effect we are suspicious of, given the low rate of scanning errors in our study. An alternative thresholding technique, adapted for polyps of varying sizes, would be a threshold based on a percentage of the ROI radius. As mentioned previous though, this would increase the complexity of data processing and potentially have the opposite effect; larger polyps would have large surrounding thresholds which then may also lead to misrepresentation of fixation in the later stages of the onscreen period.

This study is an important advance in the analysis of visual perception patterns across groups of readers who view 3D CTC images with visible polyps. It does not though demonstrate clear differences in the visual search

patterns and polyp identification characteristics of inexperienced and experienced readers.

Chapter 5: Investigating influences on reader search and performance in 3D CTC: The effect of computer-aided detection markers

Overview

- Visual gaze is attracted by computer-aided detection (CAD) marks on polyps, accelerating identification times: median ‘time to first pursuit’ was 0.48s (IQR 0.27 to 0.87s) with CAD, versus 0.58s (IQR 0.35 to 1.06s) without.
- Inexperienced readers’ gaze is affected more by CAD than experienced readers, and CAD marks hold their visual attention for longer.
- All visual search metrics demonstrated statistically significant differences when comparing reading conditions “with” and “without” CAD.
- Correct polyp identification is increased significantly by CAD (74% without CAD, 87% with CAD; $p < 0.001$)

This research has been published as;

The effect of computer-aided detection markers on visual search and reader performance during concurrent reading of CTC.

Helbren E, Fanshawe TR, Phillips P, Mallett S, Boone D, Gale A, Altman DG, Taylor SA, Manning D, Halligan S. Eur Radiol. 2015 Jun;25(6):1570-8. doi: 10.1007/s00330-014-3569-z. Epub 2015 Jan 12. PMID: 25577518

Introduction

Computer-aided detection (CAD) uses visual prompts to highlight potential abnormalities for readers of medical imaging. A wide variety of CAD systems are currently utilised across a range of modalities including mammography [126], thoracic imaging [127] and CTC [2, 128]. In CTC, CAD has been shown to improve both readers' per-patient and per-polyp sensitivity and to reduce image interpretation times in certain circumstances [1, 2, 128]. In isolation, the sensitivity of CAD systems may outperform individual readers, even when they are experienced, although careful review by a radiologist is still needed to reject false- positive prompts [129-131].

Little is known regarding how CAD systems affect visual search behaviour, a topic specifically highlighted for further investigation by the Medical Image Perception Society [132]. Studies to date have generally been laboratory based and used computer generated lesions and distractors rather than real pathology and medical images. For example, studies have focussed on how the differing visual configuration of CAD prompts influence search and have explored the potential for utilising feedback from readers' search behaviours to enhance CAD systems [101, 111, 113, 133]. By further increasing our understanding of how visual CAD prompts affect visual search, we may be able to further optimise its diagnostic performance, both in clinical practice and for reader training.

We aimed to determine the effect of CAD markers on readers' visual search behaviour and diagnostic performance in a clinically representative 3D

environment, using CTC as an example. We also aimed to investigate the relative effect of CAD on readers of differing prior experience.

Materials and Methods

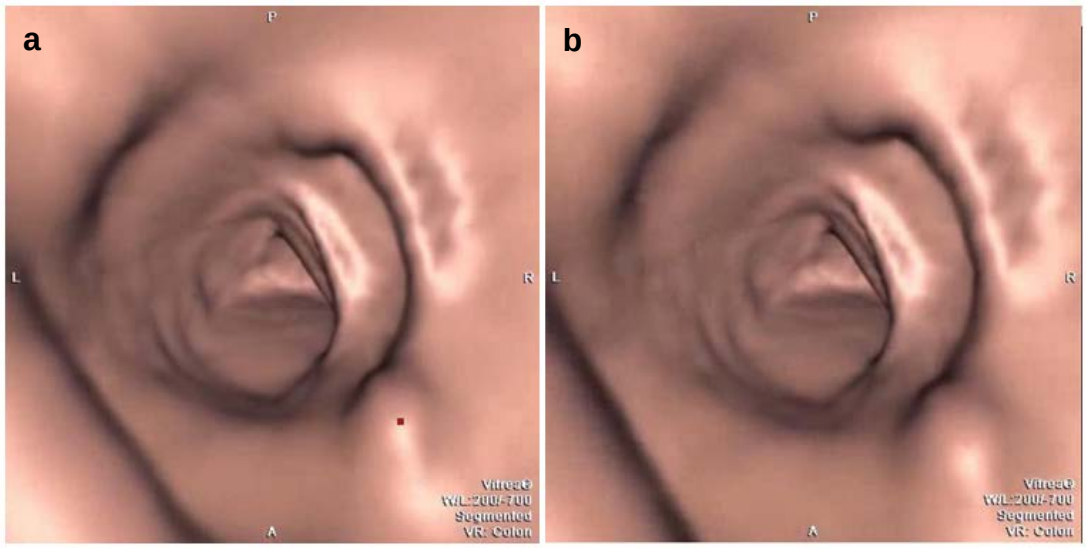
Dataset

15 CTC videos each lasting 30 seconds were recorded from the same CTC database as used in Chapters 3 and 4 [1]. The use of cases incorporated within previous studies was neither intentionally sought or avoided. For all cases, new videos were generated for use in this study (varying the orientation of the endoluminal view) by the author of this thesis. Videos were recorded from the automated, constant speed endoluminal fly-through images generated on a medical imaging workstation that utilized dedicated CTC viewing software with an integrated CAD package (Vitrea, Vital Images, Toshiba, Minnetonka, Minnesota USA). CTC videos containing a single polyp were selected on the basis of polyp size, which was at the lower boundary of clinical significance to increase the diagnostic challenge. All polyps measured between 5mm and 7mm in maximal transverse dimension via software callipers. As previously, polyps located within 5 seconds navigation of the caecal pole, rectal ampulla, or insufflation catheter were excluded as it was felt the distinctive nature of these locations might enhance recall bias. Any potential polyp that was not identified and labelled correctly by the CAD system was also excluded, as were videos with other endoluminal features that could easily be construed as polyps. Any false- positive CAD prompts were removed prior to recording.

Two videos were generated for each study; one with and one without the CAD prompt. To ensure video pairs were identical in every other way (e.g. navigation orientation, screen magnification), we used a single video generated with CAD and then our vision perception scientist (Dr Peter Phillips) generated a coloured mask matched to contextually fill colour and contrast with the surrounding colon, to hide the CAD marker in each individual frame of the without- CAD video (Figure 12).

Figure 12: With and without CAD marker screenshots

Screen shots from the same CTC video (video 2) demonstrating a polyp indicated by a CAD marker (a) and the same image after a mask had been applied to eliminate the marker (b).



Readers

42 readers participated. 17 were experienced (defined as interpretation of >200 prior CTC cases) and recruited at a subspecialty conference (2012 European Society of Gastrointestinal and Abdominal Radiology Annual Meeting). 25 were inexperienced (defined as <100 prior cases) and recruited

at either the 2012 European Society of Gastrointestinal and Abdominal Radiology Annual Meeting or at a UK teaching hospital (University College Hospital, London, England). Some participants in this study also participated in the study of reader experience (Chapter 4). In both the setting of the conference and hospital a designated quiet zone was used for tracking data collection; a quiet atrium at the conference centre and a designated room at the hospital.

Participant demographics are given in Table 8.

Table 8: Effect of CAD on performance: reader demographics

	Inexperienced (n=25)	Experienced (n=17)
Male	12 (48%)	14 (82%)
Age (years)	36.2 (6.7)*	40.4 (7.2)
Wore glasses	7 (28%)	4 (24%)
Wore contact lenses	4 (16%)	6 (35%)
Grade Staff/Consultant Trainee Radiographer Representative	 8 (32%) 15 (60%) 1 (4%) 1 (4%)	 15 (88%) 2 (12%) - -
Years' experience	7.0 (6.6)	11.5 (5.1)
Number of validated CTC cases seen previously 0 1 to 100 101 to 199 200 to 499 ≥ 500	 17 (68%) 8 (32%) - - -	 - - - 9 (53%) 8 (47%)

Table shows number (%) or mean (standard deviation) * One inexperienced reader withheld their age

Eye tracking and data collection

A Tobii X50 or X120 eye tracker (Tobii Technology AB, Danderyd, Sweden) was used as described previously in Chapter 4 (pages 75 &76) with the same five-point calibration technique prior to data collection as utilised in Chapters 3 and 4 (described page 58). A “warm-up” video was also again used to ensure data capture was successful.

Each reader viewed both video pairs (i.e. 30 individual videos). All readers were given the following instructions on screen: “You are about to see some CTC fly-throughs. Some will have CAD markers, some will not. Please click the mouse if you see a lesion you consider highly likely to represent a real polyp or cancer.” Readers were unaware of polyp prevalence. Readers were asked to hold a computer mouse immediately prior to the start of video playback so they were able to click immediately if an abnormality was seen, without needing to divert their gaze from the screen. Readers were unaware that the studies with and without CAD were paired (i.e. identical in every other way). During eye-tracking, no feedback was given to readers regarding video content or their diagnostic performance. For each reader, the order in which videos were viewed was randomised, as was whether the video was viewed first with or without the CAD marker. Videos were presented in sets of 7 or 8 at a time with an optional rest period of 5 minutes offered between sets.

Data processing

As described previously (Chapter 3, page 59) regions of interest where used to relate readers’ gaze position to the position of regions of interest on

screen. Both polyps and CAD markers were treated as separate regions of interest (ROIs). On each individual video frame on which a polyp and/or CAD marker was present, Dr Phillips applied a circular mask (i.e. ROI) that just encompassed the polyp or marker: In virtually all instances both the CAD marker and polyp were on screen simultaneously, and so the CAD marker was encompassed by the polyp ROI. However, in some circumstances the CAD marker did appear onscreen prior to the polyp (e.g. when the polyp was behind a fold). Instances of the CAD marker 'flickering' on and off the screen in short runs were managed by statistically imputing the CAD marker location in the context of its prior and subsequent visualised positions.

Eye-tracking metrics

Metrics outlined in Chapter 3 were adapted to allow analysis of the CAD marker as an ROI in the "with CAD" video set. The number of pursuits of the CAD marker location in the period prior to the polyp appearing on screen "CAD location pursuit time" was added. This comprised a) CAD location pursuit: Any pursuit of CAD marker location before the polyp appeared on screen, as a binary variable and b) CAD location pursuit time: Cumulative dwell time during period when CAD marker is visible before polyp appears on screen (expressed as proportion of time ROI on screen). Introducing a new binary metric of "immediate pursuit" to reflect the number of readers with a "time to first pursuit" of zero (since their gaze rested on the ROI immediately it appeared, as a consequence of viewing the CAD prompt that heralded the polyp's appearance and position). Differentiating correct and incorrect polyp identifications as either; "Early incorrect identification": One or more early incorrect clicks, before polyp appears on screen either with or without the

CAD marker present or a “Correct identification”: One or more correct clicks while the polyp is on screen.

Statistical analysis

Of the metrics described in Chapter 3 (pages 60 to 64) we selected the following three as primary outcome messages for this study: Time to first pursuit, Identification time span and CAD location pursuit time. Data were analysed using STATA 12.1 (StataCorp, College Station, TX). Multilevel modelling was applied, allowing for the cross-classified data structure by incorporating independent random intercepts for reader and case. Fixed effects for the presence/ absence of a CAD marker and for the inexperienced/ experienced status of the reader, as well as an interaction between these, were included. As in Chapter 4, proportional hazards models were applied for time-to-event variables, logistic models for binary variables and proportions, and Poisson models for count variables. These were reported as hazard ratio, odds ratio and rate ratio respectively, with 95% confidence interval (95% CI) and p value.

Short runs of missing longitudinal gaze data (defined as ≤50 consecutive missing observations) were again imputed based on coordinates directly before and after the times missing, with independent random noise (measurement error), with ten imputations. Cases where more than 50% of values were missing or with a run of more than 50 consecutive missing values were examined individually, being excluded from analysis if the absence of data made calculation of specific metrics unreliable. This is in contrast to the exclusion of all datasets with more than 50 consecutive

missing values, in the prior study in Chapter 4 and reflects the specific focus of this study on the interaction between CAD, polyp identification and reader gaze.

Results

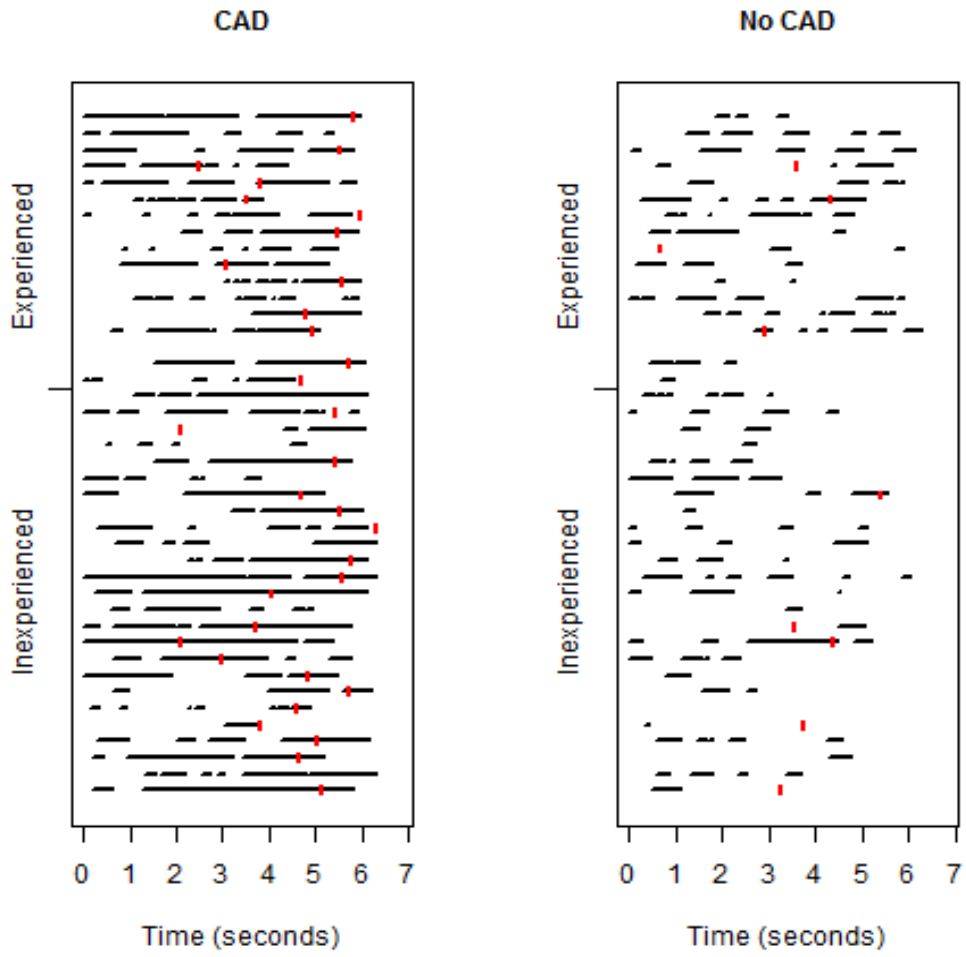
Eye tracking data was successful for all but one reader, who was excluded from analysis due to consistently poor data acquisition. Twelve other video viewings were not completed due to time pressures from competing commitments. Few specific metrics were excluded due to insufficient eye tracking data, for example 98% (1194/ 1218) of “time to first pursuit” observations were included.

Polyp identification

With CAD there was a significant increase in the overall number of correct polyp identifications across all readers; 74% without CAD, 87% with CAD. CAD increased correct identifications for both inexperienced (72% without CAD, 84% with CAD) and experienced readers (78% without CAD, 91% with CAD). However, the effect of CAD on the correct identification rate varied widely between individual videos (Figure 13).

Figure 13: ROI polyp pursuits, while the polyp is on screen, for all readers of a single video

Polyp pursuits are shown for all readers of a single video (video 4). Results separated by reader experience and presence of CAD. Black lines are pursuits with each horizontal row corresponding to an individual reader. Increased pursuit times are reflected by longer black bars: these are more frequent when CAD is used. Polyp identification time is represented by the vertical red dashes: the rate of polyp identification also increases with CAD.



Visual search

The presence of a CAD prompt drew readers' attention faster to the polyp location and led to quicker identification times (Tables 9 and 10). The percentage of eye pursuits immediately on polyps was higher with CAD (16% to 29% for inexperienced, 12% to 27% for experienced readers without and with CAD respectively; across all readers OR 2.48 [95% CI 1.84 to 3.35]). Average values for time to first pursuit, identification time span (time to click) and last assessment time span were significantly shorter in the reads with CAD. The shorter average time to first pursuit with CAD was reflected in a higher hazard ratio of 1.42 (95%CI 1.19 to 1.69) and lower median time of 0.48s (IQR 0.27 to 0.87s) compared to 0.58s (IQR 0.35 to 1.06s) without CAD. Similarly, hazard ratios were higher for the identification time span (HR 1.35, 95%CI 1.19 to 1.53) and last assessment time span (HR 1.28, 95%CI 1.12 to 1.46).

Table 9: Summary of CAD metrics.

Metrics for with-CAD and without-CAD videos, overall and by experience, across all readers.

Data are number (%) or median [inter- quartile range].

Metric	With CAD			Without CAD		
	Combined	Inexperienced	Experienced	Combined	Inexperienced	Experienced
At least one pursuit	99% (590/597)	99% (358/362)	99% (232/235)	97% (572/590)	97% (345/354)	96% (227/236)
Number of pursuits	4 [2, 6]	4 [2, 6]	4 [2, 6]	4 [2, 6]	4 [2, 6]	4 [2, 6]
Immediate pursuit (s)	28% (168/597)	29% (104/362)	27% (64/235)	14% (83/590)	16% (55/354)	12% 28/236)
Time to first pursuit (s)	0.48 [0.27,0.87]	0.48 [0.25, 0.82]	0.51 [0.29, 0.93]	0.58 [0.35, 1.06]	0.60 [0.35, 1.20]	0.57 [0.32, 0.97]
Identification time span (s)	3.00 [1.84, 5.55]	3.00 [1.87, 5.41]	3.05 [1.75, 5.65]	3.24 [1.90, 5.75]	3.23 [1.99, 5.72]	3.24 [1.84, 5.77]
Last assessment time span (s)	1.06 [0.54, 1.94]	1.15 [0.56, 2.09]	0.96 [0.48, 1.63]	1.23 [0.71, 2.19]	1.24 [0.75, 2.24]	1.21 [0.63, 2.10]
Assessment pursuit time	29% [17%, 45%]	31% [18%, 47%]	28% [14%, 41%]	25% [13%, 41%]	24% [13%, 40%]	28% [11%, 42%]
Total pursuit time	50% [35%, 63%]	56% [41%, 67%]	42% [31%, 57%]	42% [25%, 56%]	42% [24%, 57%]	41% [25%, 55%]
CAD location pursuit	72% (282/392)	74% (175/238)	69%(107/15)	44%(175/398)	44%(106/243)	45%(69/155)
CAD location pursuit time	24% (11%, 43%)	29% [12%, 50%]	19% [8%, 32%]	11% (5%, 20%)	10% [5%, 22%]	12% [5%, 19%]
Assessment pursuit rate (s)	0.78 [0.53, 1.09]	0.74 [0.52, 1.06]	0.83 [0.56, 1.12]	0.69 [0.49, 0.96]	0.66 [0.47, 0.89]	0.74 [0.52, 1.04]
Total pursuit rate (s)	0.57 [0.42, 0.77]	0.56 [0.39, 0.76]	0.63 [0.42, 0.77]	0.52 [0.37, 0.72]	0.51 [0.37, 0.70]	0.53 [0.37, 0.75]
Early incorrect identification	19% (116/622)	18% (67/367)	19% (49/255)	17% (105/619)	17% (62/365)	17% (43/254)
Correct identification	87% (542/622)	84% (309/367)	91% (233/255	74% (460/619)	72% (263/365)	78% (197/254)

CAD held readers' attention on polyps for longer, resulting in a higher average assessment pursuit rate with CAD and longer average assessment pursuit times. Figure 13 illustrates how pursuit rate and polyp identification rates varied for experienced and inexperienced readers viewing a single video both with and without CAD. The average assessment pursuit time (time spent looking at the polyp before the first mouse click) was also significantly higher with CAD (OR 1.72, 95%CI 1.45 to 2.04, $p<0.001$), increasing on average from a 25% to 29% proportion of the on-screen polyp time when a polyp was identified correctly (Table 9 & 10).

The "CAD location pursuit time" metric demonstrated that prior to the polyp coming on screen, the CAD marker held prolonged periods of gaze compared to the same screen location and time period in the without-CAD videos (OR 3.69, 95%CI 2.77 to 4.92, $p<0.001$).

Effect of experience on the visual search effect of CAD

For some metrics, significant differences were seen in the extent to which inexperienced and experienced readers were affected by CAD. Metrics demonstrating this were: time to first pursuit, CAD location pursuit time, assessment pursuit time, and total pursuit time (Table 10). Figure 14 demonstrates this effect in time to first pursuit, for which CAD was associated with a decrease in median time of 0.12 seconds for inexperienced readers but 0.06 seconds for experienced readers (Table 10).

Table 10: Eye-tracking metrics comparing inexperienced and experienced readers.

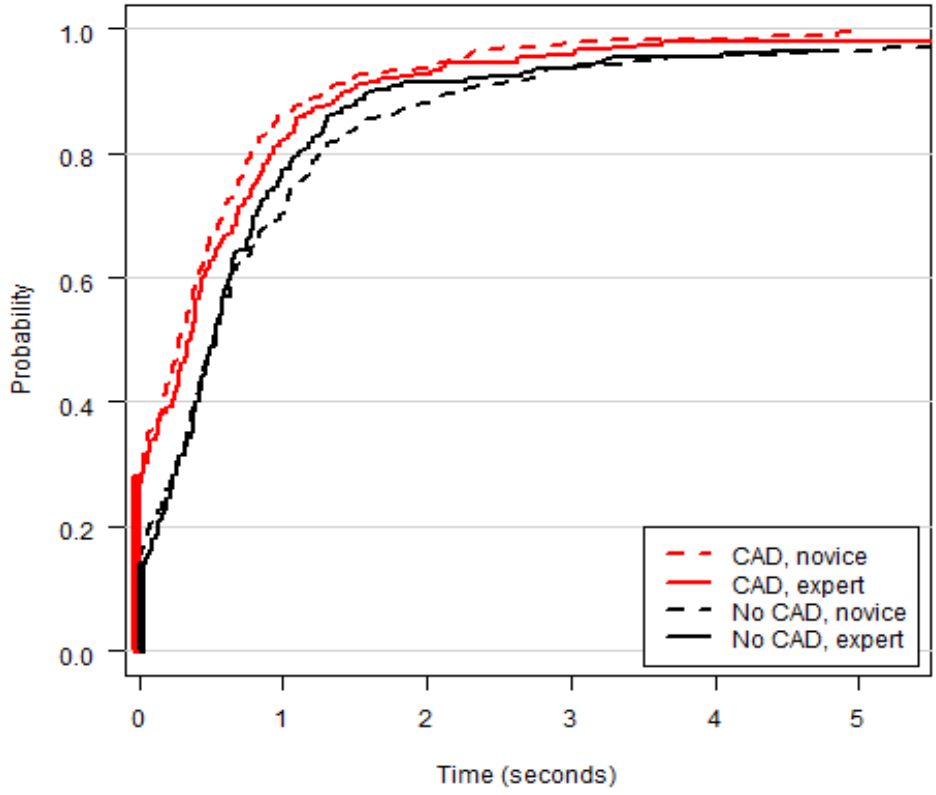
With-CAD (relative to without-CAD), experienced reader (relative to inexperienced), and the CAD: experienced reader interaction. Hazard ratio (HR), odds ratio (OR) or rate ratio (RR), 95% confidence interval and p value.

Metric	Measure	CAD		Experienced reader		CAD: experienced reader	
		Effect size [95% CI]	p	Effect size [95% CI]	p	Effect size [95% CI]	p
Immediate pursuit	OR	2.48 [1.84, 3.35]	<0.001	0.82 [0.61, 1.11]	0.20	-	-
Time to first pursuit	HR	1.42 [1.19, 1.69]	<0.001	1.29 [1.05, 1.58]	0.01	0.70 [0.54, 0.92]	-0.01
Identification time span	HR	1.35 [1.19, 1.53]	<0.001	1.17 [0.91, 1.49]	0.22	-	-
Last assessment time span	HR	1.28 [1.12, 1.46]	<0.001	1.22 [0.94, 1.58]	0.14		
Assessment pursuit time	OR	1.72 [1.45, 2.04]	<0.001	1.03 [0.73, 1.47]	0.85	0.74 [0.57, 0.98]	0.03
Total pursuit time	OR	2.00 [1.71, 2.34]	<0.001	0.94 [0.71, 1.22]	0.63	0.64 [0.50, 0.82]	<0.001
CAD location pursuit	OR	5.35 [3.65, 7.86]	<0.001	0.90 [0.56, 1.45]	0.66	-	-
CAD location pursuit time	OR	3.69 [2.77, 4.92]	<0.001	1.09 [0.75, 1.59]	0.64	0.46 [0.29, 0.73]	-0.001
Assessment pursuit rate	RR	1.18 [1.11, 1.26]	<0.001	1.11 [1.03, 1.19]	0.005	-	-
Total pursuit rate	RR	1.09 [1.03, 1.15]	0.002	1.04 [0.96, 1.13]	0.31	-	-
Early incorrect identification	OR	1.18 [0.83, 1.67]	0.36	0.95 [0.34, 2.67]	0.93	-	-
Correct identification	OR	4.21 [2.81, 6.32]	<0.001	2.22 [1.00, 4.90]	0.05	-	-

Figure 14: Time to first pursuit of polyp for readers with and without CAD

Time to first pursuit grouped by reader experience. The y axis gives the probability of a reader having their first pursuit against time since the polyp became visible on the x axis. Vertical bars at time zero represent readers with immediate pursuits.

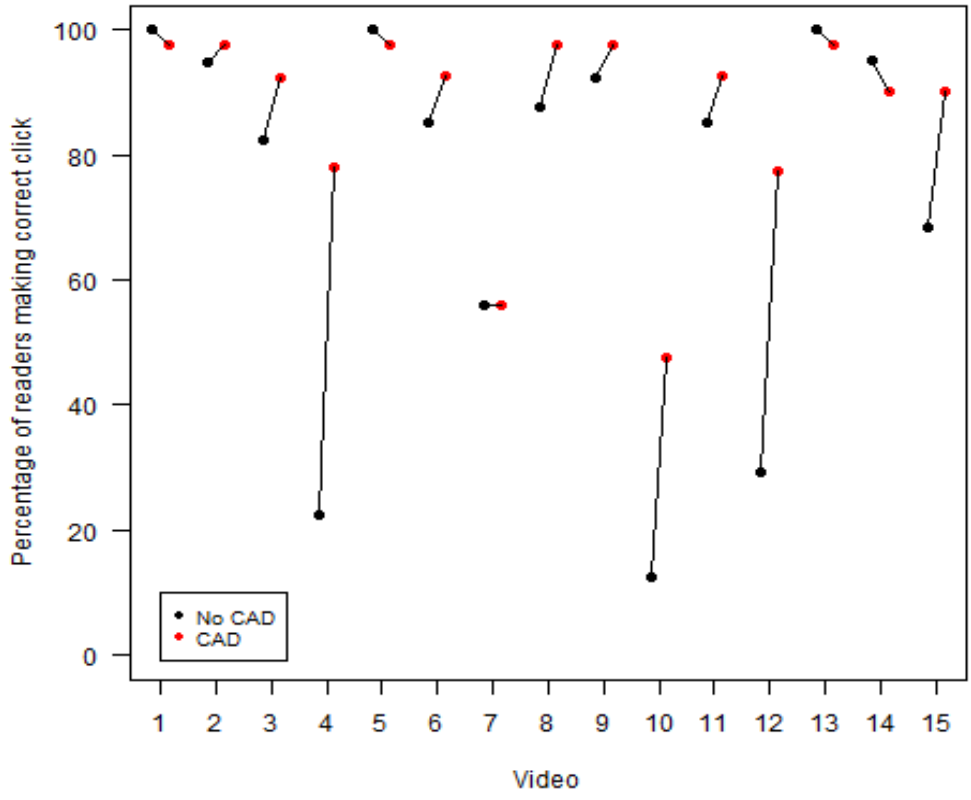
Decrease in time to first pursuit with CAD is observed in both groups but the effect is greatest for inexperienced readers (wider separation of dotted curves than solid curves).



Prior to the polyp coming on screen, the CAD marker also had a more pronounced effect as a distractor for inexperienced readers. This is exemplified by the CAD location pursuit time increasing from 10% in without-CAD novice reads to 29% with-CAD, compared to an increase from 12% to 19% respectively for experienced readers (Table 9).

Figure15: Percentage readers correct polyp identification; with and without CAD

Graph of the percentage of readers making correct polyp identifications, shown for each video. Videos with CAD are represented by red dots and those without CAD by black dots. In ten of fifteen videos, a higher percentage of readers gave a correct click when viewing with CAD (evidenced by red dots being higher than black), with the degree of increase (shown by dot separation) varying between videos.



The number of early incorrect identifications (i.e. during the period before the polyp became visible) was affected more by individual reader behaviour than the presence of CAD, with the standard deviation of reader effects being greater than the standard deviation of the metric effect. CAD did not significantly increase early incorrect identifications (17% without CAD, 19% with CAD; Table 9).

Discussion

Our results show that CAD markers significantly alter the visual search patterns for both experienced and inexperienced readers when viewing endoluminal CTC. Across all readers, all visual search metrics used to assess reader behaviour demonstrated statistically significant differences when “with” and “without” CAD reads were compared. Effects of CAD included a reduction in the time to first pursuit and total assessment time span, as well as increasing both the polyp pursuit rates and assessment pursuit times.

Prior to the polyp appearing on screen, the presence of a “herald” CAD marker greatly increased readers’ fixation on the CAD location, indicated by the higher percentage of immediate polyp pursuits for reads with-CAD. These results demonstrate how CAD markers pull visual attention before the polyp itself appears, particularly in inexperienced readers, and then hold it for prolonged periods, thereby disrupting the usual pattern of visual search.

We therefore suggest that awareness of this effect is an important feature to highlight when training radiologists to interpret CTC with CAD in the clinical setting. We hope that appreciation of the distracting effect of CAD will allow readers to focus more on active search strategies that cover other regions of the endoluminal surface. For example, it is possible that a synchronous unmarked polyp may be missed if a CAD-marked polyp is present onscreen simultaneously. Also, false-positive CAD markers could act as distractors away from true-positive polyps that are present elsewhere. Future studies should investigate these aspects. Nevertheless, CAD did exert a positive

impact on polyp identification rates in both inexperienced and experienced readers. The rate of incorrect clicks prior to the polyp appearing onscreen did not increase significantly when using CAD, suggesting that while the visual prompt may act as a distractor, it does not cause diagnostic confusion.

Of course, whether CAD is “allowed” to distract readers’ attention will depend on the paradigm within which CAD is deployed. Strictly speaking, interpretation of CAD marks simultaneously with unannotated regions of the colon should only happen during a “concurrent” reading paradigm[1, 134]. In contrast, during a “second-read” paradigm, CAD is activated only after a full unaided interpretation has been performed, in which case the reader’s gaze can jump directly from one CAD mark to another [2, 62, 128]. Therefore, the second-reader paradigm should avoid the problem of gaze being distracted by markers since CAD is not used during the initial read and only CAD markers need be inspected during the second component. However, it is unclear how readers apply the second-read paradigm in their daily clinical practice. For example, knowledge that CAD is about to be activated could accelerate the initial unaided review and so decrease vigilance. For example, prior eye tracking studies, in modalities other than CTC, have found CAD may change the way a reader utilises their time when reviewing images. For example, a laboratory study of 47 inexperienced observers asked to identify a target embedded in image noise found that CAD enhanced detection significantly but also found that unmarked targets were missed more frequently and that a lower total percentage of the search area was scrutinized than when CAD was not used at all [133]. The authors concluded that CAD does not simply combine with readers’ unaided performance in a

straightforward manner. The implication for CTC is that second-read CAD is not simply a summation of the behaviours observed in isolation with- and without CAD, a speculation that was confirmed when second-read and unassisted paradigms were compared in a prior study of CTC [2]. "Concurrent" paradigms, where CAD is activated from the start of interpretation but where interpretation is not just confined to marked areas of the colon (as occurs for "first-read" paradigms) appear to diminish sensitivity compared to second-reader implementation and so the motivation to use it is less [2]. A new paradigm, where first-read CAD is followed by a rapid unaided 2D review of the entire endoluminal surface is a variant that appears to be time- efficient and sensitive [135]. It will be interesting to see how CAD markers affect gaze under such circumstances.

We remain constrained by the need to use excerpts from CTC examinations in our eye tracking studies. The videos are much shorter than usual clinical studies and as previously, do not allow the reader to interact with navigation path and speed. However, the use of short, fixed-speed videos allowed metrics to be compared between groups of readers using multilevel modelling from a multi-reader multi-case study design. Our investigation is most pertinent to concurrent reading paradigms and does not address second-read paradigms directly, not least because methods for 3D eye-tracking are novel and relatively underdeveloped. It is expected that future automation of some of the processes necessary for image analysis will enable the use of more representative videos and other paradigms. Another limitation was the rapid repetition of duplicated case videos with and without CAD (although these were temporally separated, in random order, and

readers were not told they were seeing the same case twice). This procedure was essential to allow comparison between the two viewing environments within one viewing session, and was necessary because the most experienced readers only congregated at the single training workshop and, having done so, were then time-pressured.

In summary, by eye-tracking observers of CTC we have found that CAD markers affect multiple visual search metrics. CAD was associated with an increase in polyp identifications in both inexperienced and experienced readers. However, particularly for inexperienced readers, CAD can act as a major distractor, drawing readers' attention both before and during polyp visibility. Inexperienced readers in particular need to be aware of the potential for CAD to act as a distractor, drawing visual attention away from other regions of interest.

Chapter 6: Investigating influences on reader search and performance in 3D CTC: do prevalence expectations affect visual search and decision-making?

Overview

- Radiologists interpreted endoluminal CTC fly-throughs of the same group of 10 patients, 3 times each. Abnormality prevalence of polyps was fixed (50%) but readers were told before viewing each group that prevalence was different, either 20%, 50% or 80% in the population for which the cases were drawn
- Differences in most visual perception metrics between expected prevalence levels were not statistically significantly different.
- There was a weak tendency to look outside the central screen area at 85% prevalence and reduction in positive polyp identifications at 20% prevalence.

This research has been published in:

Do prevalence expectations affect patterns of visual search and decision-making in interpreting CT colonography endoluminal videos?

Fanshawe TR, Phillips P, Plumb A, **Helbren E**, Halligan S, Taylor SA, Gale A, Mallett S. Br J Radiol. 2016;89(1060):20150842. doi: 10.1259/bjr.20150842. Epub 2016 Feb 23. PMID: 26903391

Introduction

If we are expecting an event, we are more alert to it and therefore more likely to react more quickly and favourably when it occurs [136]. We might suspect that readers in radiology reporting are more alert to the presence of an abnormality when given an indication that the prevalence is particularly high and, conversely, be less alert to an abnormality when the chance of it occurring is thought to be low, as is the case for screening.

Interpretation of medical imaging commonly occurs in three environments: the symptomatic population, the asymptomatic/screening population and the research setting. Expected (and actual) levels of abnormality vary considerably between these groups and between medical specialties [137]. It follows that the effect of varying prevalence of abnormality on image interpretation is crucial to how diagnostic accuracy and interpretative performance might change across reporting environments.

In 2011, a systematic review [138] found only three medical imaging studies [139-141] that assessed the impact of experimentally modified prevalence on reader diagnosis. Subsequent studies have been published [142-145], but the relationship between prevalence and interpretation accuracy remains unclear. Some studies report increased false negatives or reduced diagnostic confidence at lower prevalence levels, for example, for interpretation of pulmonary arteriograms [139], mammograms [143, 146] or ankle trauma radiographs [142]. This ‘rare target’ effect has also been reported in non-clinical scenarios, such as baggage scanning [147, 148] and artificial target search experiments [149]. By contrast, in chest radiography,

the evidence for a prevalence effect on diagnostic accuracy is weaker [140, 144], although two studies that used eye tracking to monitor visual search of experienced readers suggested a possible association between increased prevalence and the duration and pattern of image scrutiny [145, 150].

Despite increasing use of CTC in routine practice, there is little research describing the effect of abnormality prevalence on diagnostic performance [138]. This is surprising because CTC is commonly applied across a wide range of expected prevalence, from asymptomatic individuals undergoing screening [151-153] to symptomatic and high-risk patients [154-156]. Establishing the presence or absence of a prevalence effect on reader attention, visual search and diagnostic performance is important both in understanding how CTC should be used in clinical practice and for designing future research studies.

The purpose of this study was to assess the effect of expected abnormality prevalence on visual search and decision-making in CTC using the techniques and metrics described in Chapter 3 and refined in Chapters 4 and 5 of this thesis.

Method and Materials

Participants and cases

13 radiologists (readers) were recruited from a UK teaching hospital (University College Hospital, London, England) over 2 days in July 2012. Some had previously participated in the studies in Chapters 4 and 5. All provided written, informed consent. Readers (6 out of 13 were males; mean

age 32 years, range 27 to 36 years) were trainees with 1 to 7 years' experience as a radiologist and at most 50 cases CTC experience.

The database of CTC studies used on prior studies in this thesis (Chapters 3-5) were again used to generate 10 new CTC endoluminal fly-through videos lasting 30 secs each. These were produced using dedicated CTC software on a medical imaging workstation (Vitrea, Vital Images Inc., MN) by the author of this thesis and then exported for viewing. Navigation speed was fixed at approximately 1.5cm s^{-1}. Five videos depicted a single colorectal polyp (true positive, 5 to 8mm maximal transverse dimension), verified by three radiologists with > 200 cases' experience [157]. As in all prior studies of this thesis, cases were excluded if they contained polyps within 5 seconds of the caecal pole, rectal ampulla or insufflation catheter, or contained other distinctive characteristics, as assessed by the author of this thesis. Polyps were on screen for between 2.4 and 11.1 seconds. The remaining five videos (true negative) were selected from different sections of the colon (which contained no polyps) in the same patient group.

The sample size was based on practical considerations: the number of readers available for recruitment and the number of cases that could comfortably be assessed comfortably in one sitting.

Data collection

The group of 10 videos was presented to each reader three times in one sitting, with an optional rest break between the groups. The order of cases within the group was randomised for each individual reader. Before viewing each group, readers were told that the videos in that group came from a

population with a known prevalence of abnormality – 20%, 50% or 80%. The ordering of the three prevalence scenarios was varied between readers using block randomisation. Readers were not told that the three groups actually contained the same 10 videos repeated three times and were therefore unaware that the true prevalence was identical (50%) and the declared 20% and 80% prevalence levels were incorrect. Information given to readers was worded as:

> "We are going to show you 3 groups of 10 videos in a random order.
>
> Each group is taken from a different population, each with a different prevalence of abnormality.
>
> Before each group we will tell you the population prevalence, either 80%, 50% or 20%."

Readers were asked to hold a computer mouse throughout and indicate with a click (polyp identification) when they saw a lesion that they considered highly likely to represent a real polyp or cancer. Readers were not required to specify polyp location and could not pause, rewind or review videos. They were not told which videos contained polyps and were given no feedback about their performance. Data collection took 20 to 30 minutes per reader.

Viewing conditions and data preparation

Reading was conducted in a quiet room with constant, ambient light. A liquid-crystal display monitor of 1280 x 1024 pixel resolution was used (SyncMaster 971P: Samsung, Suwon, Republic of Korea and Fujitsu E19-5: Fujitsu, Tokyo, Japan; 1 pixel = 0.29mm). The screen was positioned 60cm

in front of the reader. Videos measured 512 x 512 pixels (14.8 x 14.8 cm), representing a visual angle of 14.1°. Tobii X50 or X120 eye trackers, sampling at 50Hz or 60Hz respectively, were used with participant set up as described in Chapters 4 and 5. Nine-point calibration was performed prior to data collection and, as previously, readers were excluded if this could not be completed. They then viewed a supplemental warm-up video prior to data collection. They were not asked to fixate a particular point before each video.

Eye position data was prepared using ROIs as described previously in Chapter 3 and utilised in Chapters 4 and 5.

Outcome measures

Data relating to eye tracking of the polyp as a ROI was again analysed using the metrics described in Chapter 3 (pages 60 to 64).

To gain a greater understanding of the effect of the expectation of prevalence on visual search further metrics relating to off-screen time and polyp identification were incorporated. The study then reflects three aspects of reader behaviour: eye position when a polyp is on screen; eye position when no polyp is on screen; and frequency and accuracy of polyp identifications.

New metrics to reflect reader behaviour when the polyp was off screen were; “Screen coverage” Proportion of eye co-ordinates falling in to each of three regions of the screen display, “central” - a region 256 X 256 pixel square at the centre of the 512 x 512 pixel screen, “upper” - the region of the upper half of the screen outside the central area, and “lower” - the region of the lower

half of the screen outside the central area. An additional 100 pixel margin for gaze points measured outside the screen area was included with these points then incorporated into the upper or lower region as appropriate (Figure 17, page 129).

“Pursuit rate in the absence of an ROI” Number of distinct eye pursuits, divided by the total time when the polyp was off screen.

Polyp identification metrics were also expanded to include new measures to assess the potential effect of the stated abnormality prevalence on decision making;

“Total number of identifications” Number of identifications recorded over whole video

“Any correct identification” Binary indicator of whether identification occurred while the polyp was visible (a reaction time of 0.5s after the polyp left the screen was allowed)

"Any incorrect identification” Binary indicator of identifications occurring only before the polyp appeared, to prevent readers who delayed their decision after seeing a polyp being misclassified as making false-positive identification.
For true-negative videos this incorporated identifications made at any time.

Primary outcomes were; time to first pursuit of the ROI; pursuit rate in the absence of an ROI; total number of polyp identifications.

Statistical analysis

Statistical analysis used STATA v.12.1 for Windows (StataCorp, College Station, TX) and R version 3.1.1 [158] and was conducted by Dr Thomas Franshawe.

Metrics were analysed using multilevel modelling, incorporating independent random intercepts for reader and video, including prevalence level as a factor. Effects of prevalence expectation were expressed relative to the true 50% prevalence category. In a planned sensitivity analysis, to test whether results were altered by the order (first, second or third viewing) in which the prevalence categories were presented, this order was included as an additional factor variable.

Within this multilevel framework, proportional hazards, logistic and Poisson models were used, as appropriate for the data type. As most viewings had at least one missing eye position data point, short runs of missing data were imputed, based on the recorded eye co-ordinates immediately before and after, adding random measurement error. Estimates were combined using multiple imputation methods with 10 independent imputations [159]. Cases with >50% missing values or >50 consecutive missing values were examined individually by Dr Thomas Fanshawe (statistician) and Dr Andrew Plumb (radiologist) and removed if deemed likely to make the metric calculation highly unreliable.

A different approach was adopted only for pursuit rate, which has no generally agreed definition[160]. We used the number of pursuits calculated by Tobii Studio v 1.7.2 (50 pixel dispersion, 100 milliseconds minimum

threshold) throughout the period when no polyp was on screen, divided by the duration of this period. Time points when the Tobii software failed to identify whether a co-ordinate belonged to any particular pursuit were excluded, and the time denominator adjusted accordingly. As in Chapter 5 (pages 102) cases with >50% missing values of the pursuit classifier were excluded from analysis. Results are presented as point estimates with 95% confidence intervals (95% CIs) and p-values. A 5% significance level was used, unadjusted for multiple testing.

As in Chapter 4 (page 77), a proportional hazards model was used for metrics; 'Time to first pursuit' and 'Total assessment time span', as these are time-to-event variables, measuring periods of time until an identification of the polyp occurs. Only the first 'event' (pursuit of ROI, or polyp identification) was used for these variables: any events occurring subsequently, such as a duplicate identification of the same polyp or to indicate a different polyp, were discarded in the analysis of these two metrics. Cases for which no event occurred were regarded as censored at the time the polyp left the screen. Results are presented as hazard ratios.

Events that occurred at time zero, such as a reader's gaze falling within the ROI at the instant the polyp became visible, were excluded from the analysis as such events are assumed to have occurred by chance.

Logistic models were used for variables that were binary; screen coverage, which was analysed as three separate binary categories (upper, central and lower), any correct identification and any incorrect identification. The metric any incorrect identification was analysed in multiple forms: separately for all

videos, for videos with polyps and for videos without polyps. The metric assessment pursuit time, which is expressed as a proportion of the time the polyp is visible, was also analysed using a logistic model. Results are presented as odds ratios.

Poisson models were used for the three remaining metrics; assessment pursuit rate, pursuit rate and total number of identifications.

Results

Eye tracking was successful and 389 of the intended 390 intended video viewings were completed. Seven (1.8%) of these were omitted from the analysis of one or more metrics (with the exception of pursuit rate) because patterns of missing data made calculation unreliable. For pursuit rate, 37 (9.5%) of the viewings were excluded.

Table 11 summarises metrics across all readers within each prevalence scenario. Of the videos that contained a polyp, readers made at least one pursuit of the polyp for 185 of the 190 (97%) viewings with reliable data.

Table 11: Summary of metrics in expected prevalence viewed

Metric	20% prevalence	50% prevalence	80% prevalence
At least one pursuit of polyp	63/63 (100)	61/64 (95)	61/63 (97)
Immediate pursuit	5/63 (8)	4/64 (6)	10/63 (16)
Time to first pursuit (s)[a]	0.45 (0.26 to 0.65)	0.52 (0.28 to 0.82)	0.52 (0.37 to 0.95)
Total assessment time span (s)[a]	2.45 (1.33 to 5.96)	1.72 (1.00 to 3.49)	2.19 (1.15 to 5.76)
Assessment pursuit time (%)	24 (14 to 34)	21 (13 to 33)	18 (12 to 33)
Assessment pursuit rate (s^{-1})	0.59 (0.42 to 0.79)	0.56 (0.42 to 0.83)	0.69 (0.45 to 0.85)
Pursuit rate in absence of ROI (s^{-1})	2.69 (2.19 to 3.09)	2.67(2.23 to 3.02)	2.71 (2.26 to 3.11)
Screen coverage (%)			
Upper	6 (3 to 13)	7 (5 to 12)	9 (5 to 15)
Central	87 (77 to 92)	84 (77 to 90)	82 (73 to 89)
Lower	7(4 to 12)	8 (5 to 13)	8 (6 to 13)
Total number of identifications	0.75 (0.82)	0.93 (0.90)	0.97 (1.07)
Videos with polyps	1.17 (0.80)	1.38 (0.90)	1.43 (1.16)
Videos without polyps	0.34 (0.59)	0.49 (0.66)	0.51 (0.73)
Any correct identification	46/65 (71)	55/64 (86)	49/65 (75)
Any incorrect identification	39/130 (30)	48/129 (37)	51/130 (39)
Videos with polyps	21/65 (32)	22/64 (34)	25/65(38)
Videos without polyps	18/65 (28)	26/65 (40)	26/65 (40)

There were no statistically significant differences between expected prevalence levels in any metric relating to visual search while the polyp was visible (Table 11). In each prevalence scenario, readers took approximately half a second on average to direct their gaze to the ROI after the polyp appeared [hazard ratio 1.32 (95% CI 0.95 to 1.93, p=0.14) for 20% vs 50% prevalence; hazard ratio 0.95 (95% CI 0.64 to 1.40, p=0.79) for 80% vs 50% expected prevalence; Tables 12, Figure 16]. Average total assessment time span, assessment pursuit time and assessment pursuit rate were also similar in the three prevalence scenarios (Tables 11 and 12).

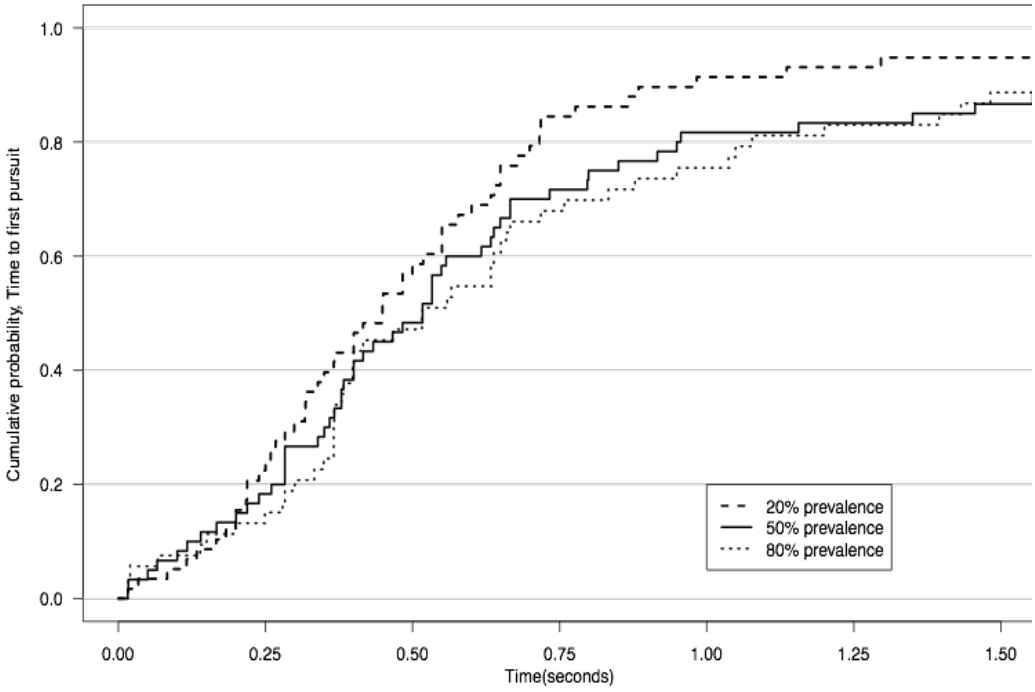
During the period when the polyp was not on the screen, the average pursuit rate was approximately 2.7 pursuits per second at each of the three prevalence levels (Table 11), with no statistically significant differences (Table 12). There was a tendency for readers' gaze to fall inside the central region of the screen less often at the 80% prevalence level than at the 50% prevalence level [odds ratio 0.82 (95% CI 0.72 to 0.95, p=0.008), Table 12, Figure 17], with a concomitant increase in the upper region. This effect, however, was small, with on average 82% of gaze points falling in the central region at 80% prevalence compared to 84% at 50% prevalence (Table 11).

Table 12: Comparison of metrics between prevalence levels

Metric	Measure	20% *vs* 50% prevalence		80% *vs* 50% prevalence	
		Effect size (95% CI)	p-value	Effect size (95% CI)	p-value
Time to first pursuit	HR	1.32 (0.95 to 1.93)	0.14	0.95 (0.64 to 1.40)	0.79
Total assessment time span	HR	0.74 (0.50 to 1.12)	0.15	0.83 (0.56 to 1.24)	0.37
Assessment pursuit time	OR	1.27 (0.87 to 1.84)	0.22	0.90 (0.62 to 1.32)	0.60
Assessment pursuit rate	RR	0.91 (0.70 to 1.18)	0.47	1.07 (0.83 to 1.37)	0.60
Pursuit rate ROI absent	RR	1.01 (0.98 to 1.05)	0.39	1.03 (1.00 to 1.07)	0.06
Screen coverage					
Upper	OR	0.93 (0.78 to 1.12)	0.45	1.28 (1.07 to 1.53)	0.007
Central	OR	1.06 (0.92 to 1.23)	0.39	0.82 (0.72 to 0.95)	0.008
Lower	OR	0.96 (0.81 to 1.13)	0.63	1.11 (0.94 to 1.31)	0.22
Total number of identifications	RR	0.81 (0.62 to 1.06)	0.12	1.04 (0.81 to 1.34)	0.75
Any correct identification	OR	0.24 (0.08 to 0.73)	0.01	0.37 (0.12 to 1.11)	0.08
Any incorrect identification	OR	0.66 (0.37 to 1.19)	0.17	1.11 (0.63 to 1.97)	0.71
Videos with polyps	OR	0.86 (0.35 to 2.11)	0.75	1.29 (0.54 to 3.10)	0.57
Videos without polyps	OR	0.53 (0.24 to 1.17)	0.11	1.00 (0.47 to 2.13)	1.00

Figure 16: Time to first pursuit in the three prevalence conditions.

Kaplan-Meier curves showing time to first pursuit in the three prevalence conditions. The vertical axis shows the proportion of viewings for which a pursuit has occurred prior to the times shown on the horizontal axis.



There were no statistically significant differences with respect to expected prevalence regarding the total number of identifications (Table 12). As expected, the average number of identifications was higher for videos that contained polyps than for those that did not (on average 1.3 vs 0.4, Table 11). The sensitivity, or probability of a polyp being correctly identified, was higher at 50% prevalence (86%) than at 20% prevalence (71%). This difference was statistically significant (p=0.01, Table 12) but the trend did not persist at the 80% prevalence level (75%). This metric was subject to an extremely high case-specific effect (Figure 18), as in three videos 1, 2 and 4 almost every reader identified the polyp at each prevalence level; the other

two videos 3 and 5, for which the polyp was superficially more difficult to identify, are therefore likely primarily responsible for the differences in rates of correct identification.

Figure 17: Screen coverage of gaze whilst polyp off screen

The division of screen area into upper, central and lower regions (dashed lines). Border outside the screen but incorporated into the data (solid line).

This example of variations in gaze across differing levels of prevalence expectation is from a single reader (Reader 11) viewing the same case (Case 3) under the three differing prevalence conditions: 20% (left panel), 50% (central panel) and 80% (right panel).

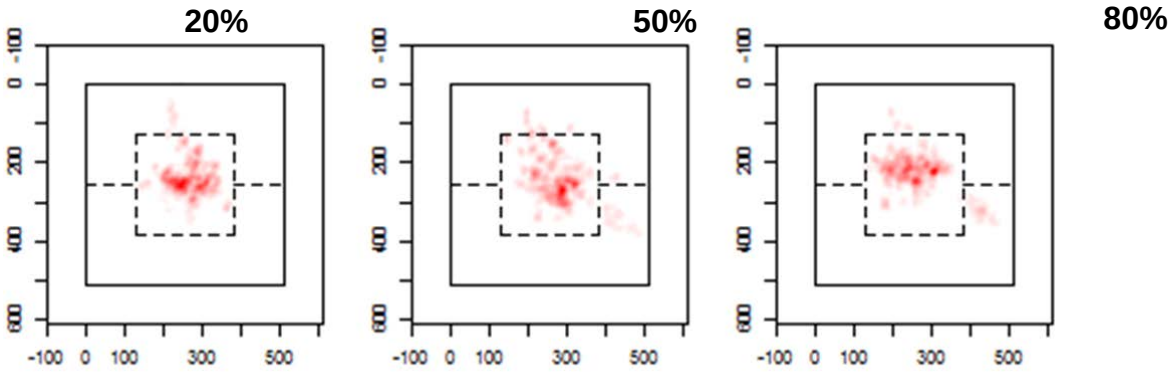
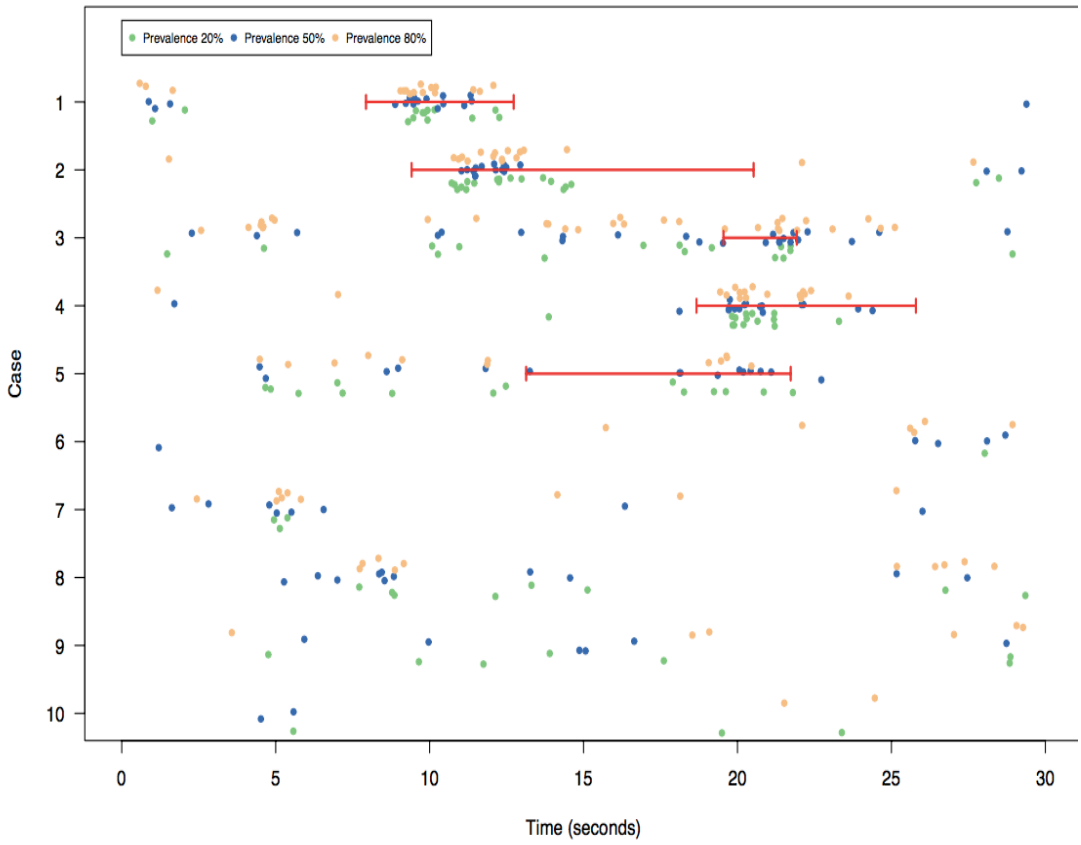


Figure 18: Time points of identifications against prevalence

Time points within each video at which polyp identifications occurred. Prevalence conditions are indicated by different colours. Cases that contain a polyp are labelled 1 to 5, and the bar indicates the period during which the polyp was visible on the screen. Cases with no polyps are labelled 6 to 10.

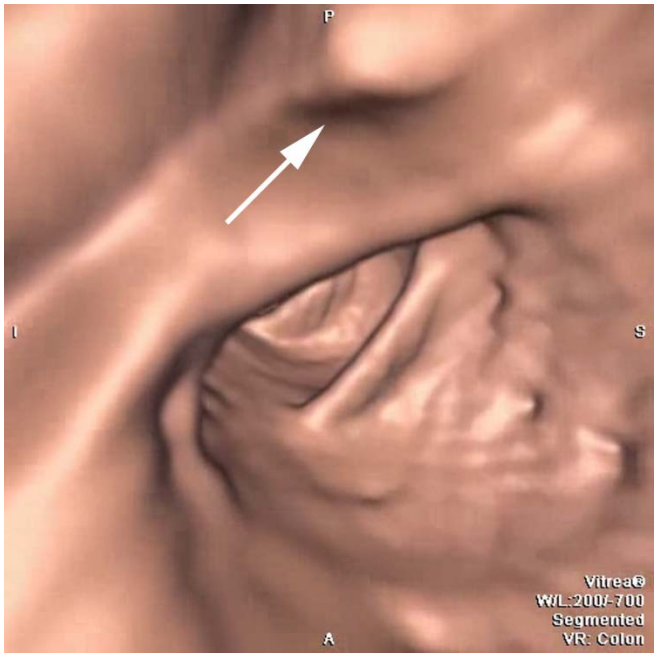


The probability of an incorrect identification (false positive) ranged from 30% at 20% prevalence to 39% at 80% prevalence; this difference was also not statistically significant (Table 12). On average, incorrect identifications occurred with similar frequency for videos that contained no polyps and for videos that contained polyps during periods when the polyp was not visible, although there was considerable variability between cases (Figure 18).

Some features were noted to commonly be identified with a mouse click by several readers' i.e. common false positives (e.g. Case 3 at 5 seconds, Figure 18 and 19).

Figure 19: Colonic feature provoking false response

Screen capture from one of the displayed videos (Case 3, at around 5 seconds) showing a feature provoking a false positive, in this case a mildly bulbous but normal fold (arrow).



In sensitivity analysis, including as an extra factor variable, the order in which the prevalence scenarios were presented did not affect the effect sizes of prevalence shown in Table 12.

Discussion

This study investigated the effect on visual search and decision making of providing readers of CTC with substantially different expectations of the likely prevalence of abnormality in the population from which cases were drawn. We did not demonstrate a strong link between prevalence expectation and the pattern of search or decision-making.

Our conclusion differs from those of several studies [143, 147-149] using scenarios other than CTC, that found increased false-negative rates at lower prevalence levels. Our study showed a statistically significant increase in the proportion of polyp identifications between 20% and 50% expected prevalence, but for three reasons these findings need to be treated cautiously. Firstly, it did not extend to the highest (80%) prevalence level, for which the proportion of false negatives was similar to that at 20%. Secondly, the effect was driven by an increased true-positive rate in just two of the five cases with polyps: a consistent increase across all cases would have provided more convincing evidence. Thirdly, this was just one of several secondary analyses performed, and so it may be a chance result.

The existence of a prevalence effect is not a universal finding in image interpretation studies. For example, Gur et al. [140] found that varying prevalence levels between 2% and 21% did not affect the diagnostic accuracy of chest radiograph assessment. Likewise, we did not find a prevalence effect for our three primary outcomes, which were chosen to represent visual search and decision-making.

We have shown previously that time to first pursuit of the polyp changes with reader experience and the presence of a computer aided detection marker (Chapters 4 and 5); in the present study, this metric was unchanged across prevalence scenarios. When no polyp was visible, we introduced new metrics to assess the readers ongoing visual search behaviours. These demonstrated that readers tended to spend more time, proportionally, looking at peripheral screen regions in the 80% prevalence condition, but this effect is small and is not supported by changes in other visual search metrics. However, the finding requires further investigation as our measure is based on a simple square at the centre of the screen area, which may not adequately capture gaze narrowing effects.

We used a common set of cases for each of the prevalence conditions to directly observe the effect of disclosing different prevalence information, as opposed to the effect of a true case-mix. Lau et al. [161] claim that the latter may have a larger effect on decision-making, but testing this was not our objective. Indeed, it would have been infeasible for readers to make an assessment of the true underlying prevalence within a realistic time frame. It is possible that some readers realised that they had viewed videos more than once, but this is unlikely to have a major effect on our findings: the order in which the prevalence conditions were presented was determined randomly and this order was not strongly associated with outcomes.

Future studies should assess further the possibility of a threshold effect in CTC. It is possible that the expected prevalence level needs to be <20% for an effect to be visible, as is usually the case in everyday clinical practice, except in very high-risk patient groups such as those examined following a

positive faecal occult blood test [156]. Evans et al [143] found a marked reduction in sensitivity for breast cancer diagnosis using mammography during screening when the prevalence was extremely low level of (0.3%). Whether a similar effect applies in CTC remains unknown. Additionally, prevalence effects may vary according to the ease of visualisation and identification of the cases chosen.

Limitations of this study include its exploratory nature, and therefore we may not have used enough cases for subtler prevalence effects to be detected. Endoluminal fly-through view was presented in automatic mode only, so readers could not adjust navigation speed as in usual practice; a limitation that we have noted previously in Chapters 4 and 5; pages 92 & 113. We were therefore unable to assess the effect of prevalence on the time the reader would spend scrutinising each video; from laboratory search experiments and some clinical studies, there is evidence that assessment time is affected by prevalence when viewing static images[150, 162]. Mouse clicks are not synonymous with definitive decisions about the presence of polyps and thus can only be regarded as proxy measures of diagnostic accuracy. Readers were not asked to identify polyp locations and so, even with eye tracking data, it is impossible to state with certainty the cause of any particular click. Readers were inexperienced in CTC, and so our findings are not directly generalisable to experienced radiologists using CTC in day-to-day clinical practice. Finally, we did not assess the effect of providing information about the spectrum of disease severity, since readers received prevalence information alone.

In summary, CTC readers were provided with different estimates of the prevalence of abnormalities from which cases were drawn, and study results did not demonstrate a strong link between prevalence information and the pattern of visual search or decision making.

Section C: Developing registration software to improve reader diagnostic performance in CTC

Chapter 7: Clinical evaluation of method for automatic co-registration of polyps at follow-up surveillance studies

Overview

- This section of the thesis investigates the benefit to the reader of applying a novel technique that facilitates the registration of the prone and supine surface location of CTC data to the follow-up of a known polyp over time in an individual patient.

This research has been published in:

CT colonography: clinical evaluation of a method for automatic coregistration of polyps at follow-up surveillance studies.

Helbren E, Roth HR, Hampshire TE, Pickhardt PJ, Taylor SA, Hawkes DJ, Halligan S.

Radiology. 2014 Nov;273(2):417-24. doi: 10.1148/radiol.14140473. Epub 2014 Jul 4.

PMID: 24991991

Introduction

Although it is generally accepted that polyps of 1 cm or larger detected at CTC should be removed, identification of smaller polyps presents a treatment dilemma, especially if those patients are at risk for colonoscopy associated adverse events or decline colonoscopy for cancer screening. Polyps can be kept under surveillance by using sequential CTC [163] with polypectomy performed only if substantial interval growth has occurred, a strategy that is believed to be both safe and cost effective [73]. Therefore, patients and clinicians may choose CTC in preference to immediate colonoscopy and polypectomy when a polyp is detected at screening CTC.

A CTC study, as described in Chapter 1 ‘Image acquisition’ page 35, incorporates the acquisition of images with the patient in two positions. Traditionally these are prone and supine but in practice frail patients may be alternatively scanned in the right and left lateral decubitus positions instead. Matching polyps on both acquisitions can be a laborious and difficult process. Distension of the colon often varies with patient position, with relative compression of the transverse colon in the prone position for example, leading to regions of suboptimal distension or complete collapse on either or both acquisitions. In such situations the matching of potential abnormalities between prone and supine studies can be challenging and prolong reporting time. Furthermore, incorrect matching may lead to diagnostic error.

Traditionally, registration between acquisitions has been based on matching locations on the colonic mucosa by their distance along the colonic

centreline. The limitation of this technique is that it does not localise to a point on the endoluminal surface, rather a point within the lumen centreline from where it is hoped the abnormality will be seen. It is also inaccurate when faced with collapsed segments since the lumen is obliterated.

When CTC is performed for polyp surveillance, the radiologist is presented with at least two CTC studies (four acquisitions): the initial study at which one or more polyps were detected and a subsequent study after an appropriate time interval. Polyps are identified on the first series of images and matched with those on images from the second series, and any growth is determined. However, matching polyps from serial CTC examinations can be challenging and time consuming, especially because such polyps are usually small.

Computer algorithms that co-register endoluminal surface locations on CTC images have been described, with the aim of facilitating and accelerating polyp matching between prone and supine acquisitions, and these are also able to account for regions of endoluminal collapse [164]. However, to our knowledge, the potential for co-registration of the endoluminal surface in temporally separate CTC investigations to facilitate polyp surveillance has not been evaluated. In our study, we aimed to evaluate the accuracy of a method for automatic co-registration of the endoluminal surfaces at CTC performed on separate occasions to facilitate identification of polyps in patients undergoing polyp surveillance.

Method

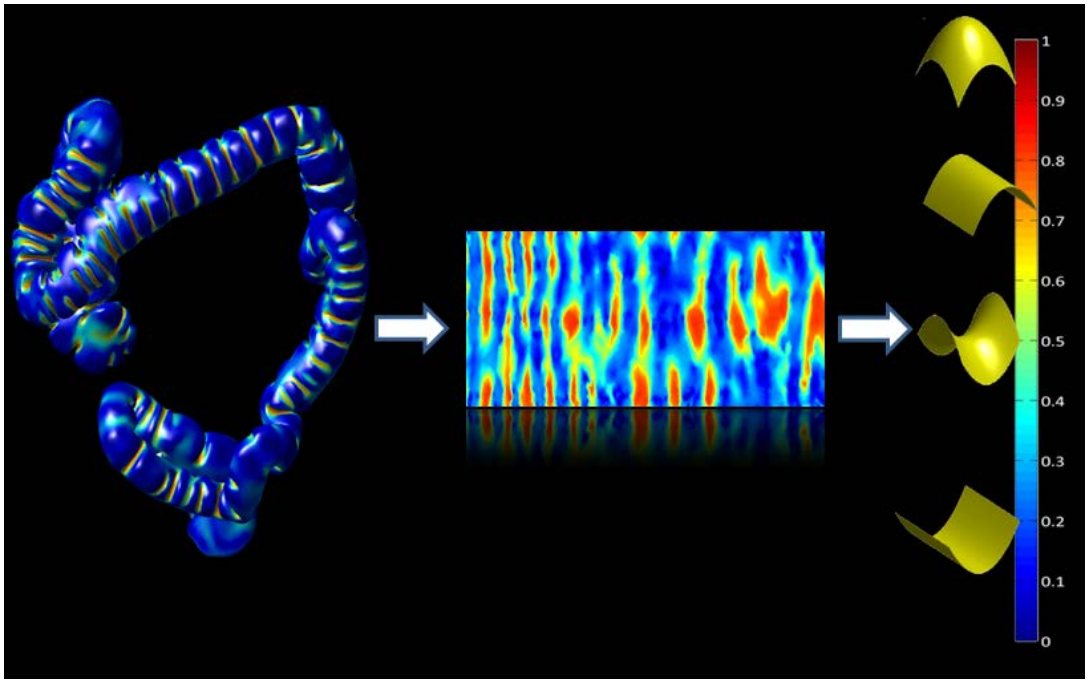
Computer algorithm:

The computer algorithm was developed by Thomas Hampshire [161] and Dr Holger Roth [162], whilst working as computer scientists in the Centre for Medical Image Computing at University College London under the supervision of Dr David Hawkes. The algorithm was refined on porcine phantoms with the assistance of Dr Darren Boone, radiologist.

The prone and supine CTC luminal segmentation images are the input to the algorithm. Surface meshes are extracted from these segmentations and it is these which are then matched. The colon is unfolded into a simpler 2D cylindrical form using the mathematic method of Ricci flow. Registration is then performed by matching landmark locations along the cylinder. This is done by using of a shape index values. The shape index gives a value to the local shape of the surface i.e. polyps would have a value close to 1 and haustral folds close to 0.75, as demonstrated in figure 20.

Figure 20: Luminal segmentation image converted by shape index

CTC luminal image converted first to a more simple 2D form and then a shape index is used to give values to the luminal contour



The extent of curvature at points along the lumen is calculated to better characterise individual points as a haustra or non-haustra. This is reflected in the curvature-based metric 'M' (Figure 21) where $M = k_1 - \gamma \| k_2 \|$

Figure 21: Curvature metric M

The metric M classifies folds by identifying long, ridge-like structures where k_1 where would be greater than 0 and k_2 close to 0.

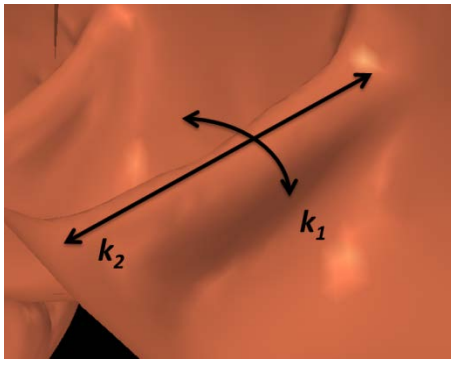


Image courtesy of Tom Hampshire

A graph cut segmentation method can then be performed over the entire surface of the mesh to define discrete haustral fold locations to match between the prone and supine studies.

This matching is achieved by a unary cost function. A virtual camera is positioned onto the interior of the colonic lumen focused on the folds to match. Depth map images are rendered at each camera location and the camera position optimised so images closely resemble one another (Figure 22).

Figure 22: Depth maps

Virtual camera images on the left are converted to depth maps on the right and camera position optimised to gain a close resemblance between folds.

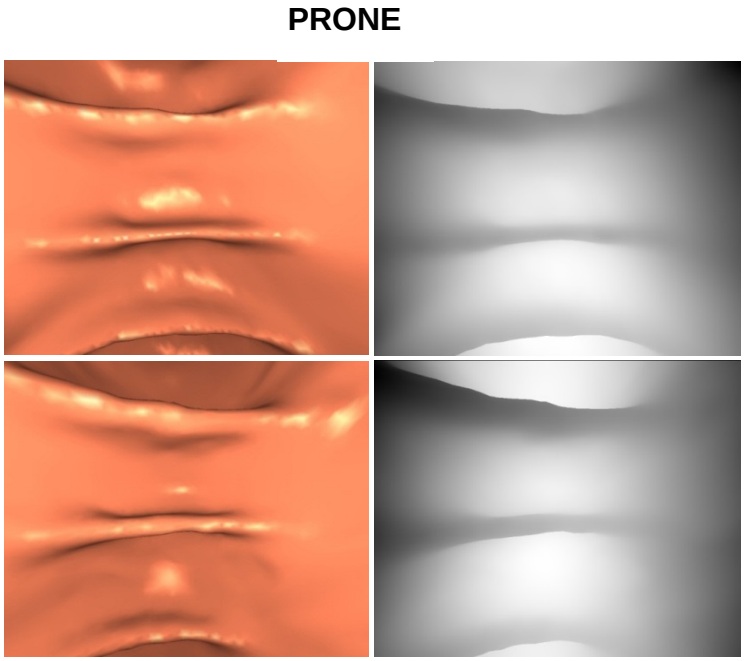


Image courtesy of Tom Hampshire

Resulting images are compared using a sum-of-squared-differences similarity metric (which gives a value directly comparing the appearance of the haustral folds).

To improve registration further, a pair-wise cost function is used to compare the similarity of pairs of haustral folds using their spatial relationship to one another (Figure 23). For each pair of neighbouring folds there is then a vector which describes the translation between each of them along the 2D cylinder. This relationship of folds should be similar in both the prone and supine positions, along with a degree of error modelled with a gaussian distribution. As the distance between folds increases so does the variance in the gaussian distribution. Sampling from this distribution gives a value for matching a pair of folds in the prone to supine acquisitions, which defines the pair-wise cost function.

Figure 23: Neighbourhood correspondence

The location of the fold labelled ‘3’ on the supine acquisition is found on the prone by comparing its relationship to the neighbouring folds (1,2,4 and 5)

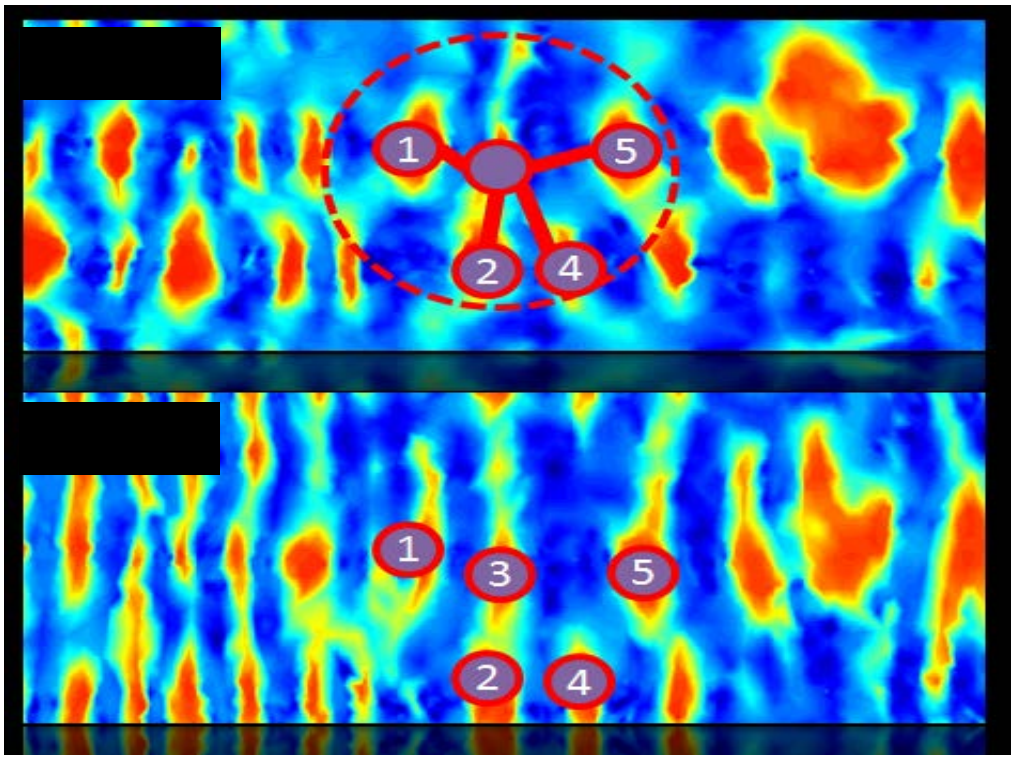


Image courtesy of Tom Hampshire

A landmark registration augments the registration. Unfolded images of the luminal surfaces are again used with the contour of the lumen highlighted. The fold matches from the two acquisitions are superimposed, along with the displacement to corresponding folds between the two lumens. A B-spline transformation grid is then used to give a smooth deformation image which gives a coarse alignment between the prone and supine (Figure 24).

Figure 24: Landmark Registration

Landmark correspondences are shown by arrows in the upper image, whilst the B-spine transformation grid is shown in the lower image.

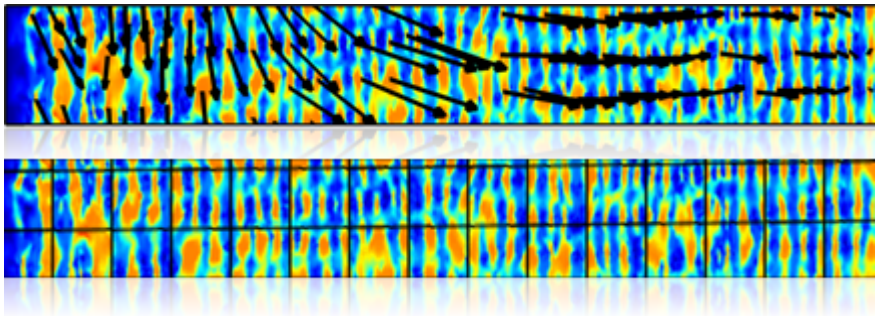


Image courtesy of Dr Holger Roth

Alignment is further improved by using an intensity based registration, using the colour intensities generated by the shape index (Figure 20). A rigid registration and then a non-rigid registration are performed to give alignment between the surface features (Figure 25).

Figure 25: Intensity based registration

Regions of intense curvature, as highlighted by the shape index, being matched between luminal surfaces.

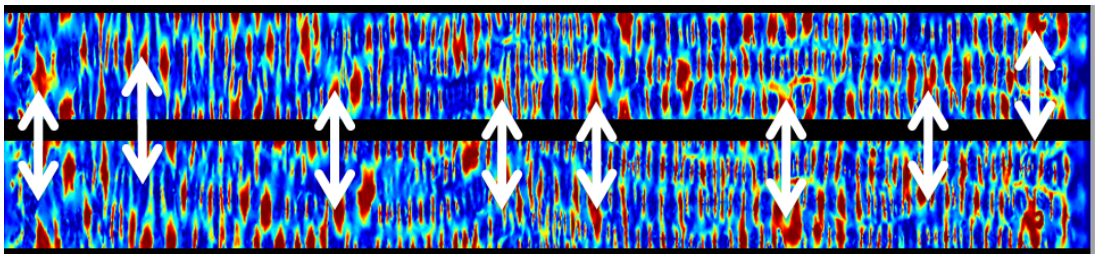


Image courtesy of Dr Holger Roth

Validation of algorithm

Clinical validation of the algorithm was undertaken in two studies [160,161];

In the first [164], the computer-assisted registration algorithm was compared to a conventional normalised distance along the colonic centreline method. A heterogeneous group of CTC examinations were used including studies with poor preparation and inadequate colonic distension. In this study a 3D polyp registration error of 19.9mm +_ 20.4 was achieved by the algorithm, compared with a 27.4mm +_ 15.1 error for the centreline method. Endoluminal review by two observers (the author of this thesis and Dr Darren Boone) showed 82% of polyp matches visible within a 120^{o} field of view by the computer assisted registration algorithm in contrast to 47% of polyp matches visible using the normalised distance along the colonic centreline method. This data then confirmed the algorithm as a new robust method of co-registration between CTC datasets.

In the second study [161], the ability to match identical endoluminal positions between the two 3D datasets acquired as part of the same CTC study was tested. The study utilised 12 well prepared, fully inflated cases, and 5 cases exhibiting one or more regions of luminal collapse. Two radiologists (the author of this thesis and Dr Andrew Plumb) and a computer scientist (Tom Hampshire) independently established a reference standard by matching haustral folds between the two patient positions; the final reference standard was achieved in consensus. In total, 1743 corresponding fold pairs were matched between the prone and supine datasets. The registration results were used to transform the positions of haustral folds from prone to supine,

and the euclidean distance between the resulting location, and that of the reference standard, was used to measure registration error. The algorithm achieved a fold matching accuracy of 96.0% and 96.1% in patient cases with and without segments of colonic collapse respectively. The mean surface registration error was 5.7mm in fully distended cases (max error 6.4mm) and 6.7mm in cases with segments of colonic collapse (max error 11.7mm). This gave an overall mean registration error of 6.0mm [165].

Case selection for the evaluation of automatic co-registration of polyps at follow-up surveillance studies

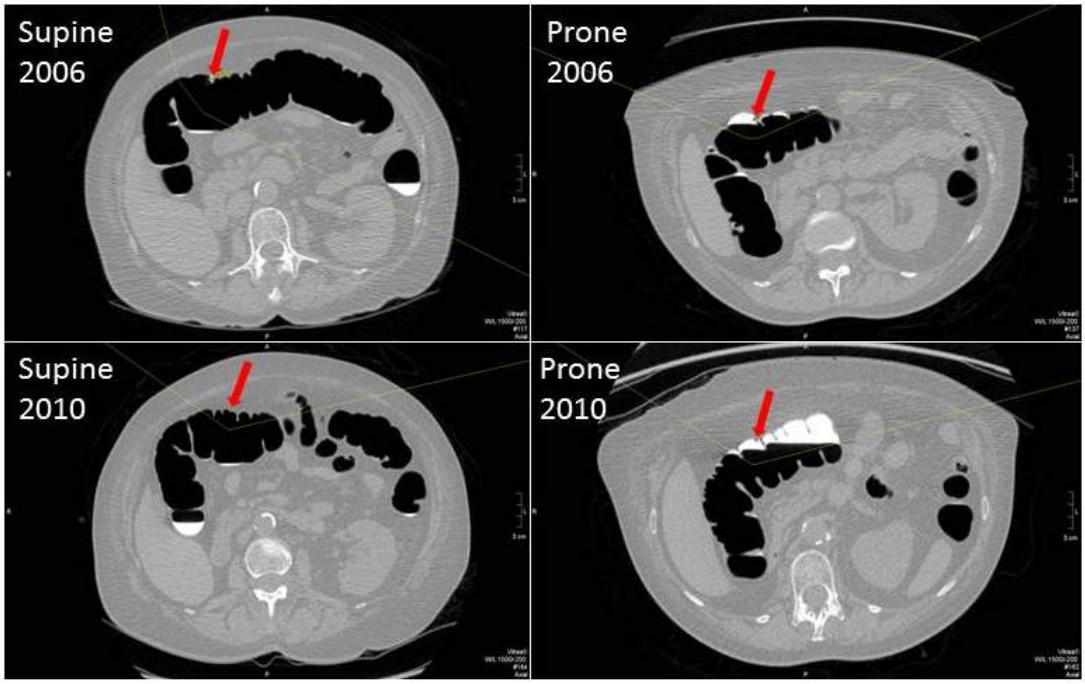
Ethical approval was held as detailed in ethics section; page 15. CTC data was taken from an institutional database of studies performed under the care of Professor Pickhardt between 2004 and 2011. All patients were asymptomatic and undergoing screening for colorectal cancer and polyps. Studies were selected by Professor Pickhardt by identifying all patients who had undergone two or more temporally separate CTC examinations during this period; all others were excluded. From the remaining studies patients who had undergone the second study for surveillance of a colorectal polyp diagnosed at the first study were selected 27 patients, with the intention to accrue 30 individual polyps or more. Selection was consecutive, and there was no attempt to bias selection in favour of patients with particularly well-prepared colons or particularly conspicuous polyps. These 27 patients each had one to three polyps, with a total of 39 polyps with a mean diameter of 6.1 mm (range, 3.6 to 9.3 mm).

For bowel preparation, the patients used a saline laxative (magnesium citrate or sodium phosphate) in conjunction with positive oral contrast material tagging (2% weight/volume barium sulfate and iodine-based diatrizoate) the evening before examination. Colonic distention was achieved with automated delivery of carbon dioxide, followed by acquisition of supine and prone low–radiation-dose multidetector CT images with 1.25mm collimation, 120 kVp, variable tube current settings, and image reconstruction with standard filtered back projection at 1mm intervals.

Studies were anonymised and study number instead assigned. Digital Information and Communication in Medicine data were transferred to compact disk, and then the studies were reviewed by the author (a radiologist with two years CTC experience) along with the original radiologic reports detailing the locations of polyps for each CTC study. The studies were uploaded to a medical imaging workstation with a CTC software package including computer aided detection (VeraLook; Vital Images). By using the polyp location information provided in the original clinical report, the author searched for the presence and location of polyps on images from both the prone and supine acquisitions (Figure 26), for initial and follow-up CTC studies.

Figure 26: Manual identification of polyp location

CTC images show manual identification of polyp location. Small transverse colon polyp (red arrows) identified at prone and supine CTC in 2006 (top images) and subsequently in 2010 (bottom images).



Studies in which the polyp was not visible on images from all acquisitions were excluded because a reference standard with which to assess the accuracy of the registration software could not be established. Where three or more acquisitions had been performed at the same examination, the author of this thesis selected the initial supine study and either the prone or decubitus view on the basis of which gave the best visualisation of the polyp; the decubitus view was selected for nine patients on images from the initial CTC study but for only one of those in the follow-up study. No patient was excluded because of poor distension or poor cleansing unless this obscured the target polyp. Collapsed segments (defined as any individual segment in

which there was complete luminal occlusion) were identified by the author and noted. Segmental polyp location was noted as cecal, ascending, transverse, descending, sigmoid, or rectal to allow any variation in the algorithm's subsequent performance to be correlated with polyp location.

Method to assess registration accuracy

To allow comparison of true and expected polyp locations, CTC studies were loaded into a Digital Imaging and Communications in Medicine viewer (ITK-SNAP, www. itksnap.org [166]). Coordinates for the point that best described the endoluminal tip of the polyp in all three planes were defined by the radiologist (the author of this thesis) who identified the coordinates of the point of maximal elevation, or endoluminal tip, of the polyp in all three planes from the surrounding endoluminal surface. This procedure provided a true reference standard for the endoluminal location for each polyp against which the accuracy of the algorithm was assessed subsequently.

We used the computerised registration algorithm which has been thoroughly described above (Computer algorithm, pages 139 – 144).

The anonymised Digital Imaging and Communications in Medicine data and initial supine polyp locations were uploaded to a nonclinical workstation by a computer scientist (Dr Holger Roth) who executed the registration algorithm. Once the registration had been processed, the reference standard coordinates of the polyp were revealed to allow any registration error to be calculated. The accuracy of endoluminal predictions given by the co-registration algorithm was tested by using two methods;

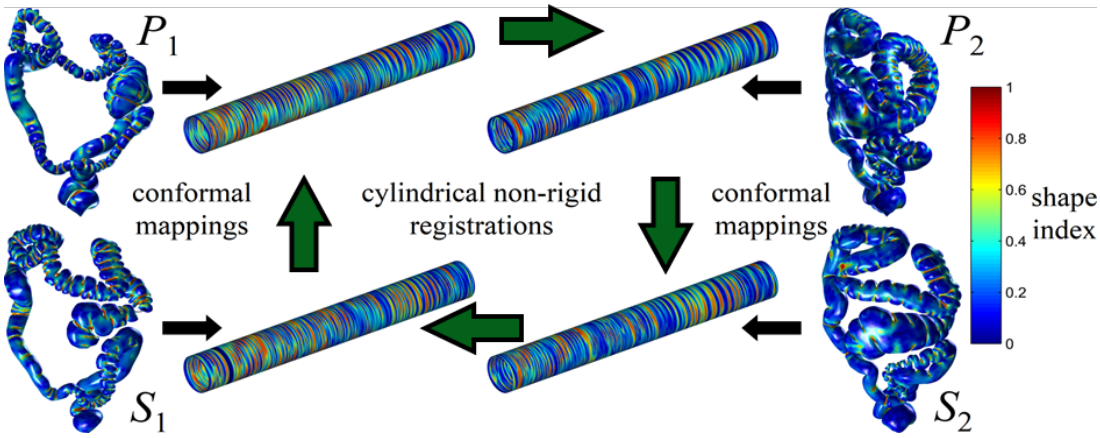
The consistency method —polyp coordinates from the initial supine acquisition were entered into the algorithm and the endoluminal polyp location expected for the corresponding initial prone acquisition was calculated by using the algorithm. These coordinates were then used to predict the endoluminal position of the polyp on the follow-up prone study images, followed by that on the corresponding supine acquisition images. The algorithm was then used with this sequence to predict the endoluminal position of the polyp back to the initial supine acquisition. This procedure allowed all points to be transformed from all data sets by computing four registrations; first supine to first prone, first prone to second prone, second prone to second supine, second supine to first supine (ie, back to first supine) (Figure 27a). Therefore, this method generated one transformation per patient. Consistency registration was assessed first, followed by longitudinal registration.

The longitudinal method —because the consistency method depends on sequential registrations, there is the potential for errors to accumulate. Therefore, the algorithm also was tested by revealing the endoluminal polyp coordinates from both of the initial acquisitions (i.e. prone and supine). A single registration from the initial supine location was then used to generate the expected polyp location on images from the follow-up supine acquisition, and the polyp location from the initial prone study was used to predict the endoluminal location of the polyp on images from the follow-up prone acquisition; then first prone to second prone and first supine to second supine (Figure 27b). Therefore, this method generated two transformations per patient.

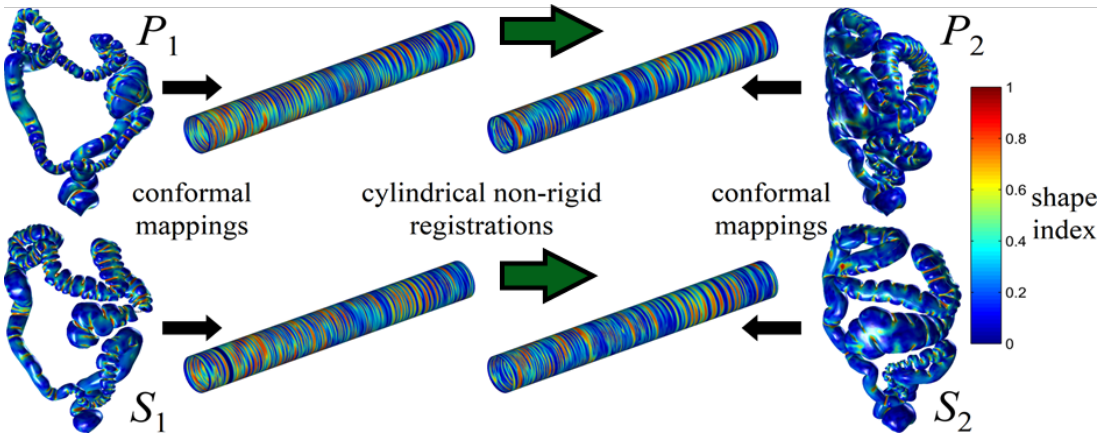
Figure 27: Methods to assess registration accuracy

Diagram shows method used to assess registration accuracy. (a) Consistency method. Error between true and expected polyp locations was calculated on all studies when all expected polyp locations were generated from initial supine polyp locations alone. (b) Longitudinal method. Error between true and expected polyp locations was calculated when polyp location was registered directly from initial supine to follow-up supine acquisition and from initial prone to follow-up prone acquisition, respectively.

a



b



Statistical analysis

To assess co-registration accuracy, we calculated the Euclidean distance in millimetres between the true and expected polyp locations (i.e. the distance between the polyp tips). Because of potential correlations between results from multiple polyps in the same patient, we performed a per-patient analysis, averaging results for each patient. Then, we calculated the overall mean co-registration error and its standard deviation. Co-registration errors in patients with and without collapsed segments were compared by using the Kruskal-Wallace test statistic. Any potential linear relationship between magnitude of co-registration error and temporal separation between examinations was investigated by using the Pearson two-tailed correlation with a p value of 0.01 to indicate a significant difference. Calculations were performed by using software (SPSS for Windows, version 21; SPSS, Chicago, Ill).

Results

After review by the author, four polyps were excluded from the study because their endoluminal locations could not be determined on images from all four CTC acquisitions because of complete segmental collapse at one or more of the acquisitions; this ultimately led to the exclusion of one patient in whom no polyp could be determined with certainty on all four acquisitions. Thus, 26 patients were included in the definitive evaluation. The mean age +/- standard deviation was 61 years (range, 51 to 79 years) at the time of the initial study and 64 years (range, 53 to 81 years) at the time of follow-up; 16 of 26 (62%) patients were men. There were a total of 35 polyps with a mean

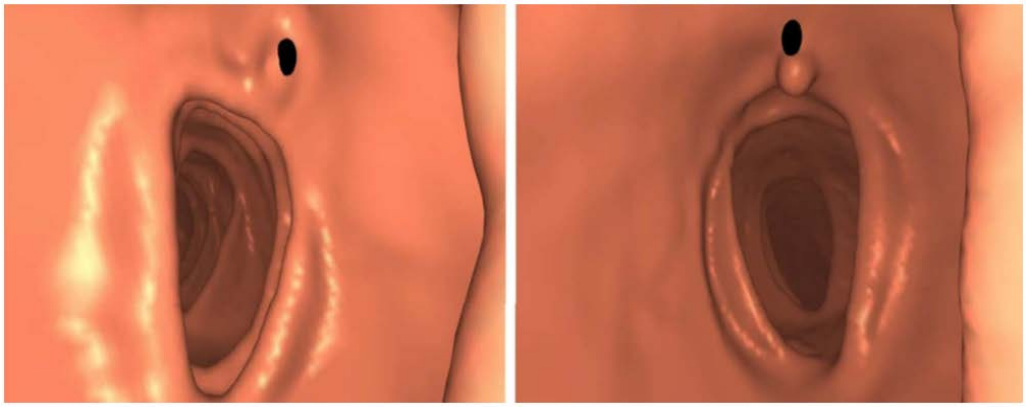
size of 6.1 mm (range, 3.6 to 9.3 mm) at the initial study and 7.7 mm (range, 4.0 to 15.3 mm) at the follow-up study. Eighteen patients had one polyp, six had two polyps, and two had three. The segmental distribution of the polyps is shown in Table 14. The mean time between studies was 898 days 6 480 (range, 266 to 1905 days). An example of the true (manually registered) polyp location compared with the expected (computerised) polyp location is shown in Figure 28.

Table 14: Co-registration polyp locations

Colonic Segment	**No. of Polyps**
Cecum	1
Ascending	4
Transverse	13
Descending	1
Sigmoid	8
Rectum	8
Total	**35**

Figure 28: Endoluminal co-registration of polyp between sequential CTC studies

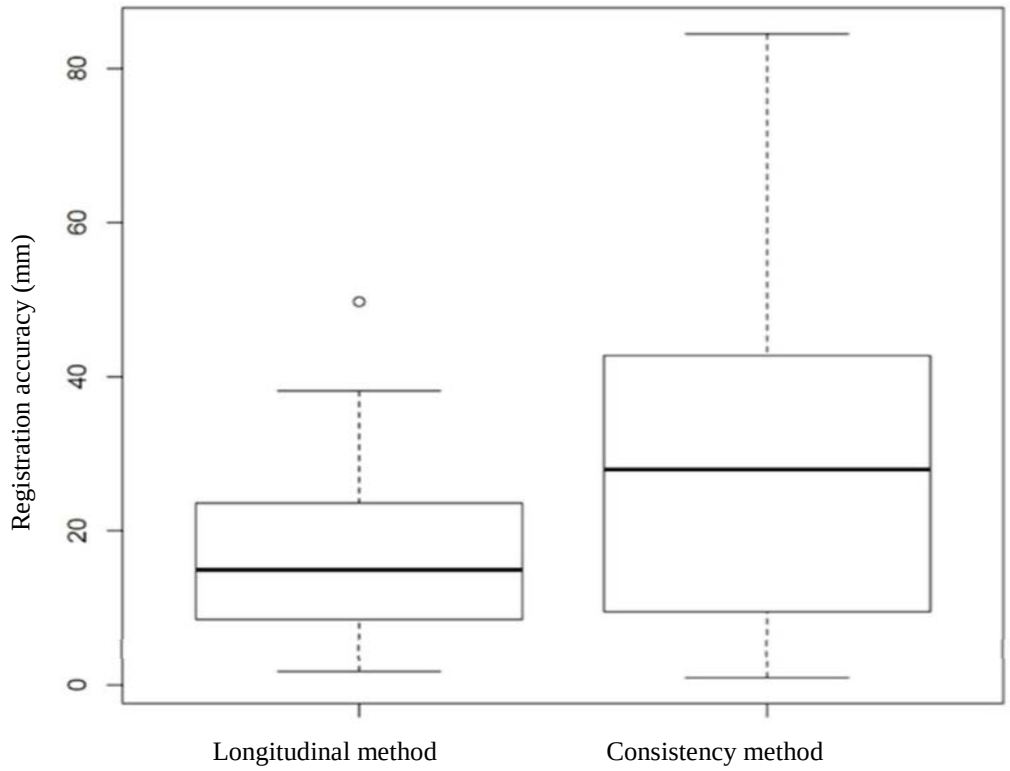
Endoluminal coregistration of polyp between sequential CTC studies. Sigmoid polyp is shown in initial supine image (left) and follow-up supine image (right). Note that polyp has increased slightly in size between studies. Black dot shows polyp location expected according to registration algorithm; in this example, coregistration was 2.9mm.



Successful polyp co-registration according to both consistency and longitudinal methods was achieved in all 26 patients for all 35 polyps (i.e. an endoluminal position for the polyp was predicted in all cases, with no technical failure). Mean Euclidean co-registration error for the consistency method was 26.9 mm +/- 20.8 (range, 0.9 to 84.5 mm) for a total of 35 consistency transformations (one for each polyp). Mean Euclidean registration error for the longitudinal method was 17.4 mm +/- 12.1 (range, 1.7 to 49.7 mm) for a total of 70 transformations (two for each polyp), Figure 29.

Figure 29: Longitudinal and consistency registration error

Graph shows data for both longitudinal and consistency registration error, demonstrating spread of errors on per-patient basis (horizontal line in box = mean, box margins = first and third quartiles, whiskers = range, with outlier as dot) and respective median errors achieved.



The consistency method incorporated 16 (45.7%) of 35 polyps from studies in which there was at least one segment of luminal collapse along the colon (none were cases in which collapse involved the segment containing the polyp, for the reasons stated in the Case Selection section). In these cases, registration error was higher, with a mean of 34.8 mm +/- 24.5 (range, 2.9 to 84.5 mm) compared with 20.5 mm +/- 17.5 (range, 0.90 to 53.3 mm) in fully distended patients, but this difference was not significant (p = 0.059).

Registration performance for temporally separate CTC acquisitions was compared with performance for acquisitions during the same examination

(i.e. first supine to first prone and second prone to second supine). We found no significant difference between these two conditions (p = 0.451 by using the Wilcoxon signed rank test with a p value of 0.01 indicative of a significant difference) with a mean registration error in the same study of 16.9 mm +/- 17.6 versus 17.4 mm +/- 12.1 between temporally separate acquisitions. There was no significant linear relationship between the magnitude of registration error and the interval between examinations: longitudinal, p = 0.105; consistency, p = 0.055.

Registration errors according to colonic segment were assessed on a per-polyp basis (Table 15). These data demonstrated that the most accurate registration for consistency registration was achieved in the descending colon, whereas the longitudinal method achieved greatest accuracy in the rectum. With consistency registration, the greatest error was observed in the ascending colon, although the transverse colon generated a very similar error (32.8 mm and 32.6 mm, respectively).

The largest errors for longitudinal registration were encountered in the sigmoid colon for both the prone and supine registrations.

Table 15: Per-polyp location co-registration results

Colonic Segment	Longitudinal Supine Error (mm)	Longitudinal Prone Error (mm)	Consistency Error (mm)
Cecum	19.7	15.3	28.8
Ascending	22.5	15.1	32.8
Transverse	19.9	21.5	32.6
Descending	17.3	13.9	6.5
Sigmoid	24.2	22.2	29.5
Rectum	8.4	12.0	15.1
Overall	18.5	16.3	26.9

Discussion

Rapid identification of polyps in patients undergoing CTC for polyp surveillance is a potentially useful application for automatic registration software. We were able to achieve automated co-registration of endoluminal polyp locations in all of the temporally separate CTC examinations that we studied. We were then able to evaluate the accuracy with which the algorithm allowed prediction of polyp location at a surveillance study from knowledge of its location at the initial CTC scan. Although the algorithm has been used previously to co-register endoluminal polyp locations in prone and supine studies performed during the same examination [164], co-registration between temporally separated examinations is more challenging because there is more likely to be greater variation in colonic residue and distension

between such examinations than between separate acquisitions at the same examination.

The lowest mean co-registration error achieved throughout the studies was 17.4mm (longitudinal method), bringing the observer immediately to within centimetres of the polyp if it was still present at the second CTC study (polyps may regress [163]). Therefore, the algorithm may aid accurate and efficient identification of known polyps in patients undergoing CTC surveillance. It may be especially helpful for locating small polyps and also for the accurate identification and registration of individual polyp locations when multiple polyps are present, a situation when conventional unassisted co-registration can be particularly challenging and time consuming for the radiologist.

The use of CTC for colon polyp surveillance has already been incorporated into management strategy at some centres, and is used for patients who have small polyps that pose no immediate risk and/or who wish to avoid colonoscopy [163]. Expansion of colorectal cancer screening programs in many countries and increasing population age and frailty suggest that surveillance with CTC will become an increasingly popular strategy for patients with low-risk polyps. Most small polyps do not present substantial risk to patients, even in the long term, because of low dysplastic grade and/ or slow growth [163]. In comparison, for many patients, the immediate risks of potential bleeding and/or perforation at colonoscopy and polypectomy are higher. An older patient with multiple co-morbidities is more at risk from colonoscopy-related adverse events than from the development of carcinomatous change in a small benign polyp. Surveillance with CTC allows

differentiation of aggressive fast-growing adenomas from slow-growing polyps, and thereby facilitates evidence-based triage of patients to appropriate treatment groups rather than exposing them to unnecessary risk.

We found that the overall standard of temporally separate co-registration, particularly that achieved by the longitudinal method, was comparable to that obtained with co-registration of prone and supine acquisitions during the same CTC examination. The registration algorithm directs the observer toward a specific location on the endoluminal surface. This contrasts with more conventional methods that attempt to co-register acquisitions by simply directing the radiologist to a normalised distance along the colonic centreline; with these methods, the true polyp location can be anywhere along the entire endoluminal circumference [167-169]. Presently, our algorithm takes approximately 2.5 minutes to segment each endoluminal surface and a further 5 minutes to co-register them together. Therefore, co-registration in two studies takes approximately 22 minutes; four segmentations and four registrations (so that both longitudinal and consistency transformations can be assessed). Segmentation and registration could be run in parallel but we did so sequentially because implementation was easier.

Our data raises the possibility that co-registration accuracy may vary according to colonic segment, because we observed better performance in some less mobile regions, with co-registration of rectal polyps superior to those in sigmoid and transverse locations. This may be due to the relatively fixed position of the rectum versus the larger deformations in shape and contour in the sigmoid and transverse colon due to their respective mesocolic attachments [170]. At the same time, we observed differing

performance of co-registration in the ascending colon (greatest consistency error) and descending colon (least consistency error), which are often less mobile than, for example, the sigmoid and transverse colon. This observation may be due to the small number of polyps in these regions (just four and one polyps, respectively). Further work is required to investigate whether more mobile segments truly present a greater challenge for polyp registration.

Our study does have limitations. We selected patients from a single centre and so have not tested algorithm performance with different bowel preparation methods and approaches to insufflation. We also restricted our sample to those patients in whom the target polyp could be visualised confidently on images from all acquisitions; this procedure was necessary so that a reference standard could be derived against which to test the algorithm. Now that we have established proof of principle, this stipulation is not strictly necessary and performance in a wider, more generalizable set of patients could be determined; in particular, performance of the algorithm to predict the location of a hidden polyp in a submerged or collapsed segment could be assessed with a larger cohort of cases. We included patients with collapsed segments as long as these segments did not contain the polyp; only one patient was excluded from our series because the polyp was in a segment of collapse. We found that mean registration error was higher for the consistency method in studies with segmental collapse, but the difference was not statistically significant; this finding may have been due to a lack of power. We now intend to apply the algorithm prospectively in clinical practice and to determine the real-world contribution that the algorithm makes toward facilitating clinical work flow and interpretation,

especially when coupled with computer-assisted detection. Future work could allow development of deformation fields used in the algorithm to estimate polyp growth or change. This may have applications both in the follow-up of polyps and in the monitoring of patients after colonic resection for endoluminal recurrence at the resection site.

In summary, we have shown that an algorithm designed to co-register endoluminal locations in prone and supine CTC acquisitions is also useful to co-register polyp locations in temporally separate CTC studies. We have shown that such software allows prediction of the endoluminal location to within a couple of centimetres of a polyp at a subsequent surveillance CTC on the basis of the coordinates from the initial study. Such software will likely facilitate the frequently time-consuming and challenging radiologic task of matching small polyps between CTC studies performed for polyp surveillance.

Section D: Conclusions and recommendations for future research

Discussion of results

The aim of this thesis was to develop our understanding and enhance the diagnostic performance of CTC through the application of novel techniques, namely eye tracking 3D moving images and sophisticated co-registration of the endoluminal surfaces. The work is prefaced by an overview of the current clinical practice of CTC and our current understanding of visual perception in medical imaging through the application of eye tracking technologies.

Chapter 3: Developing metrics that describe eye-tracking of 3D moving images

In this section we have applied and adapted traditional metrics from the 2D paradigm to the 3D environment, specifically endoluminal CTC imaging. The comprehensive collection of eye tracking metrics devised and presented within the study demonstrate how data can be extracted to reflect the interaction of visual gaze with regions of interest that move and change size through a dynamic dataset. We show how these metrics can not only define the visual search of a single reader but also a group and how they can be used to demonstrate differences between readers.

Chapter 4: Investigating influences on reader search and performance in 3D CTC: Experienced versus inexperienced readers

The metrics defined in Chapter 3 were applied to the study of the gaze of experienced and inexperienced readers with the small modification of restricting ROI at longest pursuit to pursuits occurring prior to ROI identification.

We demonstrated across all readers in all cases that experienced readers were significantly more likely to identify polyps than inexperienced readers. A significant difference in the time to first pursuit was seen between experienced and inexperienced readers with experienced readers being faster to pursue the ROI than the less experienced.

In all other metrics we were unable to establish a clear differentiation between the two groups. Nearly all readers pursed all polyps, in keeping with others subsequent findings (Further developments; page 167 &168) that recognition of pathology is a harder skill to master than scanning and fixating on pathology in medical imaging.

Chapter 5: Investigating influences on reader search and performance in 3D CTC: The effect of computer-aided detection markers

This study applied the already described metrics but with the further addition of CAD location pursuit metrics to assess gaze relating to the CAD marker prior to the polyp becoming visible on screen.

We demonstrated that the use of CAD markers significantly altered patterns of visual search in both the experienced and inexperienced readers. Across all readers a decrease in time to first pursuit and total assessment time span was seen in the 'with CAD' videos. There was also an increase in polyp pursuit rates and assessment pursuit times. A stronger effect of CAD on visual search metrics was shown in the inexperienced reader group. Ultimately, CAD led to a significant increase in the overall number of correct polyp identifications across all readers but the strength of the CAD marker, as a visual distractor, is also highlighted.

Chapter 6: Investigating influences on reader search and performance in 3D CTC: do prevalence expectations affect visual search and decision-making?

In our study on the effect of expected prevalence on eye tracking metrics in CTC we continued to evolve our metrics with the addition of measures to assess reader gaze when the ROI was not onscreen.

Changes in pre-specified prevalence levels of abnormality did not show a significant effect on reader gaze metrics. There was a weak increased tendency to look outside the central screen at a reported abnormality prevalence of 80% and a reduction in positive polyp identifications at 20% prevalence but neither were statistically significant. These findings in part may be due to the large prevalence's chosen and the small dataset numbers. Further research with larger numbers at lower prevalence levels would be of interest.

Chapter 7: Clinical evaluation of method for automatic co-registration of polyps at follow-up surveillance studies

In our final study we applied a previously validated co-registration algorithm to temporally separate CTC studies in the context of polyp follow-up. The algorithm used polyp location on the first CTC acquisition to identify the specific endoluminal location on the follow-up study (performed for polyp surveillance).

Good registration was demonstrated between studies when assessed by both longitudinal and consistency registration error. This demonstrates a further application of this technology to aid accuracy and efficiency in the follow-up of small polyps by CTC.

Further developments

Since these studies were completed there has been an increasing number of peer-reviewed articles on eye tracking in medical imaging. However, with the exception of our research group, no further studies of 3D CTC have been published. The study from our institution, from Plumb et al, looked at navigation speed and the effect on visual search and polyp detection. He found that increasing speed not only reduced visual search but also both true and false positive polyp identifications [171]. Wolfe et al does look briefly at eye tracking and search in 3D imaging in his recent studies and concludes that search errors account for more errors in 3D than 2D imaging. It is unclear though whether this is due to the increased complexity of the image

and visual information to cover or the reduction in ambiguity delivered by 3D visualisation, that then reduces decision error and therefore relatively increases search error [172].

Studies of observer expertise

Many studies have continued to investigate differences in visual search between experienced and inexperienced readers. In our study we were only able to demonstrate the time to first pursuit being linked to and shorter in experienced readers. A similar effect has also been seen in other medical imaging research and allied medical data interpretation exercises since; Wood et al demonstrated that reduced time to fixate the critical lead in ECG interpretation relates to increased accuracy of interpretation[173]. Giovanco et al, has shown that surgical orthopaedic experience and efficiency related to less overall time spent on plain film viewing [174]. Soh et al showed that a change in eye tracking metrics can be shown after relatively little tuition, demonstrating improved lesion detection, decreased time to first fixation and increased number of fixations after a single tutorial on mammography studies [175]. A study of oral and maxillofacial radiologists also showed that they made quicker first fixations than dental students when reading dental images[176].

Other studies have expanded on the relationship between search and experience. For example, experts appear to have the ability to change their search strategy when abnormality is detected as a result of their knowledge and experience. Detection only tasks will lead to less expert fixation whilst diagnostic reasoning will increase expert fixation. Gijp concludes in his

recent review of visual search and perceptual performance that expert search is characterised by a global-focal search, initially obtaining a global impression followed by a detailed focal search-to-find model [177]. This is not a new concept but new terminology has been adopted to describe the behaviours of experienced readers in viewing stacked CT images. The concept is that of two behavioural groups; the “scanners” and the “drillers” [114]. Scanners search each slice one by one whereas drillers rest their gaze in one area and scroll through the stack, or a large proportion of it, before relocating gaze on another region of the image and repeating a large volume scroll. The identification of abnormality in these cases is through motion perception, which singles out unexpected structures in the visual field. Drilling is therefore thought to be associated with expertise [177, 178]. This seems a popular idea but there has been research that appears to the contrary, with Diaz et al showing that naive observers actually viewed a higher number of slices and demonstrated more directional change than radiologists when viewing volumetric chest CT [179]. It is of course possible that eye tracking cannot yet tell us the whole story and it is clear that perception extends beyond basic fixation and pursuit, especially when dealing with more complex, high volume datasets[180].

Insights that have been gleaned via eye tracking are also proving hard to translate into useable training strategies. Kok et al observed that although experts demonstrated a more systematic inspection technique, students did not benefit from training in this [181]. Van Geel suggested that radiological appearance of disease, rather than systematic search, should be the focus of teaching [182] and Plumb also concluded that decision making, rather than

detection alone, should be the focus to improve reader performance in CTC after showing that small polyps did attract reader gaze but were none the less ignored [183]. Kelly et al showed that eye tracking metrics may change with experience level but that does not necessarily equate to improved diagnostic accuracy i.e. “move your eyes like a pro but still miss things like a novice”[184]. Gijp reflects on this in his review and highlights that the aim must be to identify search strategies which improve image perception in learners, appreciating that this may be a different thing than simply encouraging novices to replicate the search strategies of more experienced readers [177].

Image assessment is clearly a complex task and viewing strategies are likely to vary in nature in differing circumstances, for example variation of image size in CT head images led to different search strategies [185] as do variations in different forms of mammographic imaging [186]. There is also evidence that experts are more adaptable when viewing at different image speeds [187] but that radiological image interpretation is a learned skill and is task specific [188]. This last point is of special interest as it suggests that research findings cannot simply be extrapolated from other spheres and applied to the clinical radiological setting, and that caution should be exercised when interpreting data from study designs that are too laboratory based.

CAD

Since this work, the eye tracking literature related to CAD and medical imaging has focused on the possible enhancement of CAD by information

gained by eye tracking readers. Different models for this have been presented with context sensitive CAD for mammography showing potential to reduce error in reporting [189] and Computer Assisted Perception, also in mammography, increasing true positive and decreasing false positive findings [190]. In contrast, Drew and Williams conducted a series of six studies and concluded that eye-tracking feedback did not yield any reliable, sustained performance benefit but noted that for these studies search displays were used as opposed to medical imaging, and that the observers used were not described as medical professionals or trainees[191].

Prevalence

Littlefair et al has examined changes in eye position and search metrics when viewers are given differing information on the prevalence of pulmonary lesions [192]. Search was significantly longer in all images with a higher prevalence expectation level. Dwell time on true positive lesions was significantly shorter at low prevalence expectations. We were unable to show such an effect but again the markedly different search task, and complexity in analysis, needs to be emphasised between the simple reading of 2D radiographs and the 3D fly-through studies implemented in this thesis.

Wolfe et al suggests that low prevalence levels of abnormality double miss error rates in both laboratory and clinical studies. He concludes that both the perceptual decision relating to each potential abnormality fixated and the quitting threshold that discerns the time given to a target-absent response is affected by prevalence of abnormality [172]. It is therefore asserted again that further work in the medical domain, and specifically CTC, is important.

In summary, eye tracking continues to blossom as a research technique in the radiological field. We have developed a set of novel metrics for the investigation of visual search in 3D imaging and demonstrated how these can be applied to compare reader performance. Since the completion of the work in this thesis, our institution has gone on to further utilise these techniques in further studies. It is hoped that in time the value of this will also be recognised by others who can then also apply it the investigation of visual search in the 3D environment.

Sometimes the results from eye tracking studies are disappointing and uncertainty persists regarding the actual contribution of this seemingly useful technology for “real world” radiological practice [191]. Some believe that eye tracking simply “confirms the obvious”. However, that conclusion may reflect a lack of refinement around the methods used to process and analyse the data obtained [193]. Clearly there are difficulties in preserving clinical validity during the academic study of visual search, especially within complex, multifactorial environments. As such, widely available software is likely required to aid eye tracking research for stack reporting CT and 3D imaging so that other authors can replicate our methods and extend the work described in this thesis into other arenas [178].

Conclusion

Despite the rapid development and widespread adoption of CTC which utilises the 3D endoluminal ‘virtual colonoscopy’ display, our understanding of the unique visual task required for interpretation remains limited. Visual perception research is still struggling to deal with analysis of digital, scrollable, stacked cross-sectional imaging and the translation into the virtual world of image viewing from within the body marks a further significant advance in complexity for this field of research. If visual perception and specifically eye tracking is to remain relevant to clinical practice and training, then it needs to move further into clinical studies, such as those described within this thesis, exploring more relevant but necessarily complex environments. We have shown that it is possible and valuable to continue to develop novel techniques in 3D CTC imaging, our eye tracking studies assessing and aiding our understanding of performance and our co-registration techniques aiding interpretation. The author hopes that the worked described in this thesis will act as a springboard for further research within this domain and allow us to better utilise CTC in the early detection of neoplastic lesions.

Appendices

Publications arising from this thesis

Towards a framework for analysis of eye-tracking studies in the three dimensional environment: a study of visual search by experienced readers of endoluminal CT colonography.

Helbren E, Halligan S, Phillips P, Boone D, Fanshawe TR, Taylor SA, Manning D, Gale A, Altman DG, Mallett S.

Br J Radiol. 2014 May;87(1037):20130614. doi: 10.1259/bjr.20130614. Epub 2014 Feb 20. PMID: 24689842

The effect of computer-aided detection markers on visual search and reader performance during concurrent reading of CT colonography.

Helbren E, Fanshawe TR, Phillips P, Mallett S, Boone D, Gale A, Altman DG, Taylor SA, Manning D, Halligan S.

Eur Radiol. 2015 Jun;25(6):1570-8. doi: 10.1007/s00330-014-3569-z. Epub 2015 Jan 12. PMID: 25577518

Tracking eye gaze during interpretation of endoluminal three-dimensional CT colonography: visual perception of experienced and inexperienced readers.

Mallett S, Phillips P, Fanshawe TR, **Helbren E**, Boone D, Gale A, Taylor SA, Manning D, Altman DG, Halligan S.

Radiology. 2014 Dec;273(3):783-92. doi: 10.1148/radiol.14132896. Epub 2014 Jul 15. PMID: 25028782

Do prevalence expectations affect patterns of visual search and decision-making in interpreting CT colonography endoluminal videos?

Fanshawe TR, Phillips P, Plumb A, **Helbren E**, Halligan S, Taylor SA, Gale A, Mallett S.

Br J Radiol. 2016;89(1060):20150842. doi: 10.1259/bjr.20150842. Epub 2016 Feb 23. PMID: 26903391

CT colonography: clinical evaluation of a method for automatic coregistration of polyps at follow-up surveillance studies.

Helbren E, Roth HR, Hampshire TE, Pickhardt PJ, Taylor SA, Hawkes DJ, Halligan S.

Radiology. 2014 Nov;273(2):417-24. doi: 10.1148/radiol.14140473. Epub 2014 Jul 4.

PMID: 24991991

REFERENCES

1. Halligan, S., et al., Computed tomographic colonography: assessment of radiologist performance with and without computer-aided detection. Gastroenterology, 2006. **131**(6): p. 1690-9.
2. Halligan, S., et al., Incremental benefit of computer-aided detection when used as a second and concurrent reader of CT colonographic data: multiobserver study. Radiology, 2011. **258**(2): p. 469-76.
3. (NCIN), N.C.I.N. *Colorectal Canncer Survival by Stage*. 2009 2018]; Available from: http://www.ncin.org.uk/cancer_type_and_topic_specific_work/cancer_type_specific_work/colorectal_cancer/.
4. Winawer, S.J., et al., Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. N Engl J Med, 1993. **329**(27): p. 1977-81.
5. Mulder, S.A., et al., Prevalence and prognosis of synchronous colorectal cancer: a Dutch population-based study. Cancer Epidemiol, 2011. **35**(5): p. 442-7.
6. Heald, R.J. and H.J. Bussey, Clinical experiences at St. Mark's Hospital with multiple synchronous cancers of the colon and rectum. Dis Colon Rectum, 1975. **18**(1): p. 6-10.
7. Hewitson, P., et al., Cochrane systematic review of colorectal cancer screening using the fecal occult blood test (hemoccult): an update. Am J Gastroenterol, 2008. **103**(6): p. 1541-9.
8. Steele, R., et al., Use of faecal occult blood tests in symptomatic patients. BMJ, 2015. **351**: p. h4256.
9. Network, S.I.G. *Diagnosis and management of colorectal cancer*. 2016; Available from: http://www.sign.ac.uk/assets/sign126.pdf.
10. Song, L.L. and Y.M. Li, Current noninvasive tests for colorectal cancer screening: An overview of colorectal cancer screening tests. World J Gastrointest Oncol, 2016. **8**(11): p. 793-800.
11. Lin, J.S., et al., Screening for Colorectal Cancer: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. JAMA, 2016. **315**(23): p. 2576-94.
12. Rex, D.K., et al., Colorectal Cancer Screening: Recommendations for Physicians and Patients from the U.S. Multi-Society Task Force on Colorectal Cancer. Am J Gastroenterol, 2017. **112**(7): p. 1016-1030.
13. Adler, A., et al., Improving compliance to colorectal cancer screening using blood and stool based tests in patients refusing screening colonoscopy in Germany. BMC Gastroenterol, 2014. **14**: p. 183.
14. Su, B.B., H. Shi, and J. Wan, Role of serum carcinoembryonic antigen in the detection of colorectal cancer before and after surgical resection. World J Gastroenterol, 2012. **18**(17): p. 2121-6.
15. Atkin, W., et al., Computed tomographic colonography versus colonoscopy for investigation of patients with symptoms suggestive of colorectal cancer (SIGGAR): a multicentre randomised trial. Lancet, 2013. **381**(9873): p. 1194-202.

16. Gavin, D.R., et al., The national colonoscopy audit: a nationwide assessment of the quality and safety of colonoscopy in the UK. Gut, 2013. **62**(2): p. 242-9.
17. Rees, C.J., et al., UK key performance indicators and quality assurance standards for colonoscopy. Gut, 2016. **65**(12): p. 1923-1929.
18. Committee, A.S.o.P., et al., *Complications of colonoscopy.* Gastrointest Endosc, 2011. **74**(4): p. 745-52.
19. Shah, H.A., et al., Factors associated with incomplete colonoscopy: a population-based study. Gastroenterology, 2007. **132**(7): p. 2297-303.
20. Atkin, W., et al., Is whole-colon investigation by colonoscopy, computerised tomography colonography or barium enema necessary for all patients with colorectal cancer symptoms, and for which patients would flexible sigmoidoscopy suffice? A retrospective cohort study. Health Technol Assess, 2017. **21**(66): p. 1-80.
21. Choices, N. *Bowel scope screening*. Bowel cancer screening 2015 2018]; Available from: https://www.nhs.uk/conditions/bowel-cancer-screening/bowel-scope-screening/#more-information-about-bowel-scope-screening.
22. Painter, J., et al., Depth of insertion at flexible sigmoidoscopy: implications for colorectal cancer screening and instrument design. Endoscopy, 1999. **31**(3): p. 227-31.
23. Halligan, S., et al., Computed tomographic colonography versus barium enema for diagnosis of colorectal cancer or large polyps in symptomatic patients (SIGGAR): a multicentre randomised trial. Lancet, 2013. **381**(9873): p. 1185-93.
24. Plumb, A.A., et al., Use of CT colonography in the English Bowel Cancer Screening Programme. Gut, 2014. **63**(6): p. 964-73.
25. Spada, C., et al., Clinical indications for computed tomographic colonography: European Society of Gastrointestinal Endoscopy (ESGE) and European Society of Gastrointestinal and Abdominal Radiology (ESGAR) Guideline. Eur Radiol, 2015. **25**(2): p. 331-45.
26. Taylor, S.A., et al., European Society of Gastrointestinal and Abdominal Radiology (ESGAR): consensus statement on CT colonography. Eur Radiol, 2007. **17**(2): p. 575-9.
27. Levin, B., et al., Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. Gastroenterology, 2008. **134**(5): p. 1570-95.
28. McFarland, E.G., et al., ACR Colon Cancer Committee white paper: status of CT colonography 2009. J Am Coll Radiol, 2009. **6**(11): p. 756-772 e4.
29. Neri, E., et al., The second ESGAR consensus statement on CT colonography. Eur Radiol, 2013. **23**(3): p. 720-9.
30. Liedenbaum, M.H., et al., Low-fiber diet in limited bowel preparation for CT colonography: Influence on image quality and patient acceptance. AJR Am J Roentgenol, 2010. **195**(1): p. W31-7.
31. Lefere, P.A., et al., Dietary fecal tagging as a cleansing method before CT colonography: initial results polyp detection and patient acceptance. Radiology, 2002. **224**(2): p. 393-403.
32. Nagata, K., et al., Full-laxative versus minimum-laxative fecal-tagging CT colonography using 64-detector row CT: prospective blinded comparison of diagnostic performance, tagging quality, and patient acceptance. Acad Radiol, 2009. **16**(7): p. 780-9.

33. Iannaccone, R., et al., Computed tomographic colonography without cathartic preparation for the detection of colorectal polyps. Gastroenterology, 2004. **127**(5): p. 1300-11.
34. Zueco Zueco, C., et al., CT colonography without cathartic preparation: positive predictive value and patient experience in clinical practice. Eur Radiol, 2012. **22**(6): p. 1195-204.
35. Lefere, P., et al., CT colonography after fecal tagging with a reduced cathartic cleansing and a reduced volume of barium. AJR Am J Roentgenol, 2005. **184**(6): p. 1836-42.
36. Taylor, S.A., et al., CT colonography: optimisation, diagnostic performance and patient acceptability of reduced-laxative regimens using barium-based faecal tagging. Eur Radiol, 2008. **18**(1): p. 32-42.
37. Fletcher, J.G., et al., Noncathartic CT colonography: Image quality assessment and performance and in a screening cohort. AJR Am J Roentgenol, 2013. **201**(4): p. 787-94.
38. Cai, W., et al., Informatics in radiology: dual-energy electronic cleansing for fecal-tagging CT colonography. Radiographics, 2013. **33**(3): p. 891-912.
39. Burling, D., et al., Automated insufflation of carbon dioxide for MDCT colonography: distension and patient experience compared with manual insufflation. AJR Am J Roentgenol, 2006. **186**(1): p. 96-103.
40. Shinners, T.J., et al., Patient-controlled room air insufflation versus automated carbon dioxide delivery for CT colonography. AJR Am J Roentgenol, 2006. **186**(6): p. 1491-6.
41. Taylor, S.A., et al., Optimizing colonic distention for multi-detector row CT colonography: effect of hyoscine butylbromide and rectal balloon catheter. Radiology, 2003. **229**(1): p. 99-108.
42. Rogalla, P., et al., Spasmolysis at CT colonography: butyl scopolamine versus glucagon. Radiology, 2005. **236**(1): p. 184-8.
43. Agency, M.a.H.p.R. *Hyoscine butylbromide (Buscopan) injection: risk of serious adverse effects inpatients with underlying cardiac disease*. volume 10 issue 7 2017 February 2017: 1; Available from: https://www.gov.uk/drug-safety-update/hyoscine-butylbromide-buscopan-injection-risk-of-serious-adverse-effects-in-patients-with-underlying-cardiac-disease.
44. Pickhardt, P.J., Incidence of colonic perforation at CT colonography: review of existing data and implications for screening of asymptomatic adults. Radiology, 2006. **239**(2): p. 313-6.
45. Sosna, J., et al., Colonic perforation at CT colonography: assessment of risk in a multicenter large cohort. Radiology, 2006. **239**(2): p. 457-63.
46. Lara, L.F., et al., The safety of same-day CT colonography following incomplete colonoscopy with polypectomy. United European Gastroenterol J, 2015. **3**(4): p. 358-63.
47. Kim, S.Y., et al., Automated carbon dioxide insufflation for CT colonography: effectiveness of colonic distention in cancer patients with severe luminal narrowing. AJR Am J Roentgenol, 2008. **190**(3): p. 698-706.
48. Gryspeerdt, S.S., et al., Supine/left decubitus scanning: a valuable alternative to supine/prone scanning in CT colonography. Eur Radiol, 2004. **14**(5): p. 768-77.
49. Buchach, C.M., D.H. Kim, and P.J. Pickhardt, Performing an additional decubitus series at CT colonography. Abdom Imaging, 2011. **36**(5): p. 538-44.

50. Liedenbaum, M.H., H.W. Venema, and J. Stoker, Radiation dose in CT colonography--trends in time and differences between daily practice and screening protocols. Eur Radiol, 2008. **18**(10): p. 2222-30.
51. Berrington de Gonzalez, A., K.P. Kim, and J. Yee, CT colonography: perforation rates and potential radiation risks. Gastrointest Endosc Clin N Am, 2010. **20**(2): p. 279-91.
52. Mainenti, P.P., et al., Accuracy of single phase contrast enhanced multidetector CT colonography in the preoperative staging of colo-rectal cancer. Eur J Radiol, 2006. **60**(3): p. 453-9.
53. Graser, A., et al., Dose reduction and image quality in MDCT colonography using tube current modulation. AJR Am J Roentgenol, 2006. **187**(3): p. 695-701.
54. Pickhardt, P.J., et al., Primary 2D versus primary 3D polyp detection at screening CT colonography. AJR Am J Roentgenol, 2007. **189**(6): p. 1451-6.
55. Plumb, A.A., et al., CT colonography in the English Bowel Cancer Screening Programme: national survey of current practice. Clin Radiol, 2013. **68**(5): p. 479-87.
56. Schmidt, S.A., et al., 3D detection of colonic polyps by CT colonography: accuracy, pitfalls, and solutions by adjunct 2D workup. Clin Radiol, 2015. **70**(10): p. 1144-51.
57. Programme, N.B.C.S. *Guidelines for the use of imaging in the NHS Bowel Cancer Screening Programme*. 2012; Second Edition:[Available from: http://www.bcsp.nhs.uk/files/nhsbcsp05.pdf.
58. Radiologists, T.R.C.o. *Guidelines on the use of CT colonography for suspected colorectal cancer*. 2014; Available from: https://www.rcr.ac.uk/system/files/publication/field_publication_files/BFCR(14)9_COLON.pdf.
59. Liedenbaum, M.H., et al., Evaluation of a standardized CT colonography training program for novice readers. Radiology, 2011. **258**(2): p. 477-87.
60. Boone, D., et al., CT colonography: who attends training? A survey of participants at educational workshops. Clin Radiol, 2011. **66**(6): p. 510-6.
61. Taylor, S.A., et al., CT colonography: investigation of the optimum reader paradigm by using computer-aided detection software. Radiology, 2008. **246**(2): p. 463-71.
62. Regge, D., et al., Efficacy of computer-aided detection as a second reader for 6-9-mm lesions at CT colonography: multicenter prospective trial. Radiology, 2013. **266**(1): p. 168-76.
63. Boone, D., et al., Assessment of the Incremental Benefit of Computer-Aided Detection (CAD) for Interpretation of CT Colonography by Experienced and Inexperienced Readers. PLoS One, 2015. **10**(9): p. e0136624.
64. Bedenne, L., et al., Adenoma--carcinoma sequence or "de novo" carcinogenesis? A study of adenomatous remnants in a population-based series of large bowel cancers. Cancer, 1992. **69**(4): p. 883-8.
65. Fidler, J.L., et al., Detection of flat lesions in the colon with CT colonography. Abdom Imaging, 2002. **27**(3): p. 292-300.
66. Pickhardt, P.J., et al., Flat colorectal lesions in asymptomatic adults: implications for screening with CT virtual colonoscopy. AJR Am J Roentgenol, 2004. **183**(5): p. 1343-7.
67. Fidler, J. and C. Johnson, Flat polyps of the colon: accuracy of detection by CT colonography and histologic significance. Abdom Imaging, 2009. **34**(2): p. 157-71.

68. Bonnington, S.N. and M.D. Rutter, Surveillance of colonic polyps: Are we getting it right? World J Gastroenterol, 2016. **22**(6): p. 1925-34.
69. Butterly, L.F., et al., Prevalence of clinically important histology in small adenomas. Clin Gastroenterol Hepatol, 2006. **4**(3): p. 343-8.
70. Odom, S.R., et al., The rate of adenocarcinoma in endoscopically removed colorectal polyps. Am Surg, 2005. **71**(12): p. 1024-6.
71. Stryker, S.J., et al., Natural history of untreated colonic polyps. Gastroenterology, 1987. **93**(5): p. 1009-13.
72. Barancin, C., et al., Prospective blinded comparison of polyp size on computed tomography colonography and endoscopic colonoscopy. Clin Gastroenterol Hepatol, 2011. **9**(5): p. 443-5.
73. Pickhardt, P.J., et al., Clinical management of small (6- to 9-mm) polyps detected at screening CT colonography: a cost-effectiveness analysis. AJR Am J Roentgenol, 2008. **191**(5): p. 1509-16.
74. Zalis, M.E., et al., CT colonography reporting and data system: a consensus proposal. Radiology, 2005. **236**(1): p. 3-9.
75. Cairns, S.R., et al., Guidelines for colorectal cancer screening and surveillance in moderate and high risk groups (update from 2002). Gut, 2010. **59**(5): p. 666-89.
76. Pickhardt, P.J., et al., Colorectal cancer: CT colonography and colonoscopy for detection--systematic review and meta-analysis. Radiology, 2011. **259**(2): p. 393-405.
77. Dominitz, J.A. and B. Spiegel, Editorial: On the Quality of Quality Metrics: Rethinking What Defines a Good Colonoscopy. Am J Gastroenterol, 2016. **111**(5): p. 730-2.
78. Taylor, S.A., et al., Acceptance by patients of multidetector CT colonography compared with barium enema examinations, flexible sigmoidoscopy, and colonoscopy. AJR Am J Roentgenol, 2003. **181**(4): p. 913-21.
79. van Gelder, R.E., et al., CT colonography and colonoscopy: assessment of patient preference in a 5-week follow-up study. Radiology, 2004. **233**(2): p. 328-37.
80. Park, S.H., et al., CT colonography for detection and characterisation of synchronous proximal colonic lesions in patients with stenosing colorectal cancer. Gut, 2012. **61**(12): p. 1716-22.
81. Kim, D.H., et al., CT colonography versus colonoscopy for the detection of advanced neoplasia. N Engl J Med, 2007. **357**(14): p. 1403-12.
82. Johnson, C.D., et al., Accuracy of CT colonography for detection of large adenomas and cancers. N Engl J Med, 2008. **359**(12): p. 1207-17.
83. Graser, A., et al., Comparison of CT colonography, colonoscopy, sigmoidoscopy and faecal occult blood tests for the detection of advanced adenoma in an average risk population. Gut, 2009. **58**(2): p. 241-8.
84. Liedenbaum, M.H., et al., Using CT colonography as a triage technique after a positive faecal occult blood test in colorectal cancer screening. Gut, 2009. **58**(9): p. 1242-9.
85. Regge, D., et al., Diagnostic accuracy of computed tomographic colonography for the detection of advanced neoplasia in individuals at increased risk of colorectal cancer. JAMA, 2009. **301**(23): p. 2453-61.
86. Halligan, S., CT colonography for investigation of patients with symptoms potentially suggestive of colorectal cancer: a review of the UK SIGGAR trials. Br J Radiol, 2013. **86**(1026): p. 20130137.

87. Pickhardt, P.J., et al., Colorectal and extracolonic cancers detected at screening CT colonography in 10,286 asymptomatic adults. Radiology, 2010. **255**(1): p. 83-8.
88. Rayner, K., et al., Eye movements and visual encoding during scene perception. Psychol Sci, 2009. **20**(1): p. 6-10.
89. Johansson, R., et al., Eye movements during scene recollection have a functional role, but they are not reinstatements of those produced during encoding. J Exp Psychol Hum Percept Perform, 2012. **38**(5): p. 1289-314.
90. Packer, E., et al., Visual analytics for spatial clustering: using a heuristic approach for guided exploration. IEEE Trans Vis Comput Graph, 2013. **19**(12): p. 2179-88.
91. Kundel, H.L., C.F. Nodine, and E.A. Krupinski, Searching for lung nodules. Visual dwell indicates locations of false-positive and false-negative decisions. Invest Radiol, 1989. **24**(6): p. 472-8.
92. Krupinski, E.A., Visual scanning patterns of radiologists searching mammograms. Acad Radiol, 1996. **3**(2): p. 137-44.
93. Hu, C.H., et al., Searching for bone fractures: a comparison with pulmonary nodule search. Acad Radiol, 1994. **1**(1): p. 25-32.
94. Carmody, D.P., C.F. Nodine, and H.L. Kundel, An analysis of perceptual and cognitive factors in radiographic interpretation. Perception, 1980. **9**(3): p. 339-44.
95. Kundel, H.L., C.F. Nodine, and D. Carmody, Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. Invest Radiol, 1978. **13**(3): p. 175-81.
96. Berbaum, K.S., et al., Cause of satisfaction of search effects in contrast studies of the abdomen. Acad Radiol, 1996. **3**(10): p. 815-26.
97. Niimi, R., et al., Eye-tracking device comparisons of three methods of magnetic resonance image series displays. J Digit Imaging, 1997. **10**(4): p. 147-51.
98. Timberg, P., et al., Investigation of viewing procedures for interpretation of breast tomosynthesis image volumes: a detection-task study with eye tracking. Eur Radiol, 2013. **23**(4): p. 997-1005.
99. Ellis, S.M., et al., Thin-section CT of the lungs: eye-tracking analysis of the visual approach to reading tiled and stacked display formats. Eur J Radiol, 2006. **59**(2): p. 257-64.
100. Mello-Thoms, C., The problem of image interpretation in mammography: effects of lesion conspicuity on the visual search strategy of radiologists. Br J Radiol, 2006. **79 Spec No 2**: p. S111-6.
101. Kundel, H.L., C.F. Nodine, and E.A. Krupinski, Computer-displayed eye position as a visual aid to pulmonary nodule interpretation. Invest Radiol, 1990. **25**(8): p. 890-6.
102. Kundel, H.L., et al., Holistic component of image perception in mammogram interpretation: gaze-tracking study. Radiology, 2007. **242**(2): p. 396-402.
103. Kundel, H.L., et al., Using gaze-tracking data and mixture distribution analysis to support a holistic model for the detection of cancers on mammograms. Acad Radiol, 2008. **15**(7): p. 881-6.
104. Donovan, T., D. Litchfield, and T.J. Crawford, Editorial: Medical Image Perception: How Much Do We Understand It? Front Psychol, 2017. **8**: p. 2072.
105. Barrett, J.R., et al., Unobtrusive evaluation of mammographer's eye movements during diagnosis of mammograms. Radiographics, 1996. **16**(1): p. 167-72.

106. Anderson, B. and C.R. Shyu, Studying visual behaviors from multiple eye tracking features across levels of information representation. AMIA Annu Symp Proc, 2011. **2011**: p. 72-9.
107. Arora, R., H.L. Kundel, and C.A. Beam, Perceptually based FROC analysis. Acad Radiol, 2005. **12**(12): p. 1567-74.
108. Chakraborty, D.P., H.J. Yoon, and C. Mello-Thoms, Application of threshold-bias independent analysis to eye-tracking and FROC data. Acad Radiol, 2012. **19**(12): p. 1474-83.
109. Phillips, P., et al., Method for tracking eye gaze during interpretation of endoluminal 3D CT colonography: technical description and proposed metrics for analysis. Radiology, 2013. **267**(3): p. 924-31.
110. Tall, M., et al., Accuracy of a remote eye tracker for radiologic observer studies: effects of calibration and recording environment. Acad Radiol, 2012. **19**(2): p. 196-202.
111. Krupinski, E.A., C.F. Nodine, and H.L. Kundel, A perceptually based method for enhancing pulmonary nodule recognition. Invest Radiol, 1993. **28**(4): p. 289-94.
112. Carmody, D.P., et al., Machine classification of dental images with visual search. Acad Radiol, 2001. **8**(12): p. 1239-46.
113. Tourassi, G.D., et al., Exploring the potential of context-sensitive CADe in screening mammography. Med Phys, 2010. **37**(11): p. 5728-36.
114. Drew, T., et al., Scanners and drillers: characterizing expert visual search through volumetric images. J Vis, 2013. **13**(10).
115. Krupinski, E.A. and R.M. Nishikawa, Comparison of eye position versus computer identified microcalcification clusters on mammograms. Med Phys, 1997. **24**(1): p. 17-23.
116. Krupinski, E.A., Visual search of mammographic images: influence of lesion subtlety. Acad Radiol, 2005. **12**(8): p. 965-9.
117. Matsumoto, H., et al., Where do neurologists look when viewing brain CT images? An eye-tracking study involving stroke cases. PLoS One, 2011. **6**(12): p. e28928.
118. Leong, J.J., et al., Visual search behaviour in skeletal radiographs: a cross-specialty study. Clin Radiol, 2007. **62**(11): p. 1069-77.
119. Chakraborty, D., H.J. Yoon, and C. Mello-Thoms, Spatial localization accuracy of radiologists in free-response studies: Inferring perceptual FROC curves from mark-rating data. Acad Radiol, 2007. **14**(1): p. 4-18.
120. White, I.R., P. Royston, and A.M. Wood, Multiple imputation using chained equations: Issues and guidance for practice. Stat Med, 2011. **30**(4): p. 377-99.
121. Phillips, P., et al., Method for Tracking Eye Gaze during Interpretation of Endoluminal 3D CT Colonography: Technical Description and Proposed Metrics for Analysis. Radiology, 2013.
122. Nodine, C.F. and H.L. Kundel, Using eye movements to study visual search and to improve tumor detection. Radiographics, 1987. **7**(6): p. 1241-50.
123. Nodine, C.F., et al., Time course of perception and decision making during mammographic interpretation. AJR Am J Roentgenol, 2002. **179**(4): p. 917-23.
124. Lund, P.J., et al., Comparison of conventional and computed radiography: assessment of image quality and reader performance in skeletal extremity trauma. Acad Radiol, 1997. **4**(8): p. 570-6.
125. Nodine, C.F., et al., Nature of expertise in searching mammograms for breast masses. Acad Radiol, 1996. **3**(12): p. 1000-6.

126. Gilbert, F.J., et al., Single reading with computer-aided detection for screening mammography. N Engl J Med, 2008. **359**(16): p. 1675-84.
127. Mazzone, P.J., et al., Lung cancer screening with computer aided detection chest radiography: design and results of a randomized, controlled trial. PLoS One, 2013. **8**(3): p. e59650.
128. Dachman, A.H., et al., Effect of computer-aided detection for CT colonography in a multireader, multicase trial. Radiology, 2010. **256**(3): p. 827-35.
129. Taylor, S.A., et al., Computer-assisted reader software versus expert reviewers for polyp detection on CT colonography. AJR Am J Roentgenol, 2006. **186**(3): p. 696-702.
130. Taylor, S.A., et al., CT colonography and computer-aided detection: effect of false-positive results on reader specificity and reading efficiency in a low-prevalence screening population. Radiology, 2008. **247**(1): p. 133-40.
131. Lawrence, E.M., et al., Colorectal polyps: stand-alone performance of computer-aided detection in a large asymptomatic screening population. Radiology, 2010. **256**(3): p. 791-8.
132. Krupinski, E.A. and K.S. Berbaum, The Medical Image Perception Society update on key issues for image perception research. Radiology, 2009. **253**(1): p. 230-3.
133. Drew, T., C. Cunningham, and J.M. Wolfe, When and why might a computer-aided detection (CAD) system interfere with visual search? An eye-tracking study. Acad Radiol, 2012. **19**(10): p. 1260-7.
134. Hock, D., et al., Virtual dissection CT colonography: evaluation of learning curves and reading times with and without computer-aided detection. Radiology, 2008. **248**(3): p. 860-8.
135. Iussich, G., et al., Computer-aided detection for computed tomographic colonography screening: a prospective comparison of a double-reading paradigm with first-reader computer-aided detection against second-reader computer-aided detection. Invest Radiol, 2014. **49**(3): p. 173-82.
136. Deese, J., Some problems in the theory of vigilance. Psych Rev, 1955. **62**(5): p. 359-368.
137. Kundel, H.L., Disease prevalence and radiological decision making. Invest Radiol, 1982. **17**: p. 107-109.
138. Boone, D., et al., Systematic review: bias in imaging studies - the effect of manipulating clinical context, recall bias and reporting intensity. Eur Radiol, 2012. **22**(3): p. 495-505.
139. Egglin, T.K.P. and A.R. Feinstein, Context bias: A problem in diagnostic radiology. JAMA, 1996. **276**: p. 1752-1755.
140. Gur, D., et al., Prevalence effect in a laboratory environment. Radiology, 2003. **228**(1): p. 10-14.
141. Gur, D., et al., The prevalence effect in a laboratory environment: Changing the confidence ratings. Acad Radiol, 2007. **14**(1): p. 49-53.
142. Pusic, M.V., et al., Prevalence of abnormal cases in an image bank affects the learning of radiograph interpretation. Med Educ, 2012. **46**: p. 289-298.
143. Evans, K.K., R.L. Birdwell, and J.M. Wolfe, If you don't find it often, you often don't find it: Why some cancers are missed in breast cancer screening. PLoS ONE, 2013. **8**(5): p. e64366.
144. Nocum, D.J., et al., The effect of abnormality-prevalence expectation on naïve observer performance and visual search. Radiography, 2013. **19**(3): p. 196-199.

145. Reed, W.M., et al., Can prevalence expectations drive radiologists' behavior? Acad Radiol, 2014. **21**: p. 450-456.
146. Gur, D., et al., The "Laboratory" Effect: Comparing radiologists' performance and variability during prospective clinical and laboratory mammography interpretations. Radiology, 2008. **249**(1): p. 47-53.
147. Wolfe, J.M., T.S. Horowitz, and N.M. Kenner, Rare items often missed in visual searches. Nature, 2005. **435**: p. 439-440.
148. Wolfe, J.M., et al., Low target prevalence is a stubborn source of errors in visual search tasks. J Exp Psychol Gen, 2007. **136**(4): p. 623-38.
149. Rich, A.N., et al., Why do we miss rare targets? Exploring the boundaries of the low prevalence effect. J Vis, 2008. **8**(15): p. 15 1-17.
150. Reed, W.M., et al., The effect of abnormality-prevalence expectation on expert observer performance and visual search. Radiology, 2011. **258**(3): p. 938-943.
151. Pickhardt, P.J., et al., Computed Tomographic Virtual Colonoscopy to Screen for Colorectal Neoplasia in Asymptomatic Adults. New England Journal of Medicine, 2003. **349**(23): p. 2191-2200.
152. Johnson, C.D., et al., Accuracy of CT Colonography for Detection of Large Adenomas and Cancers. New England Journal of Medicine, 2008. **359**(12): p. 1207-1217.
153. Stoop, E.M., et al., Participation and yield of colonoscopy versus non-cathartic CT colonography in population-based screening for colorectal cancer: a randomised controlled trial. Lancet Oncology, 2011. **13**(1): p. 55-64.
154. Atkin, W., et al., Computed tomographic colonography versus colonoscopy for investigation of patients with symptoms suggestive of colorectal cancer (SIGGAR): a multicentre randomised trial. Lancet, 2013. **381**(9873): p. 1194-1202.
155. Regge, D., et al., Diagnostic Accuracy of Computed Tomographic Colonography for the Detection of Advanced Neoplasia in Individuals at Increased Risk of Colorectal Cancer. JAMA, 2009. **301**(23): p. 2453-2461.
156. Liedenbaum, M.H., et al., Using CT colonography as a triage technique after a positive faecal occult blood test in colorectal cancer screening. Gut, 2009. **58**: p. 1242-1249.
157. Halligan, S., et al., Incremental benefit of Computer-Aided Detection when used as a second and concurrent reader of CT Colonographic data: Multiobserver study. Radiology, 2011. **258**(2): p. 469-476.
158. R Core Team, *R: A language and environment for statistical computing*. 2012, R Foundation for Statistical Computing: Vienna, Austria.
159. Barnard, J. and D.B. Rubin, Small-sample degrees of freedom with multiple imputation. Biometrika, 1999. **86**(4): p. 948-955.
160. Salvucci, D. and J. Goldberg. Identifying fixations and saccades in eye-tracking protocols. in Proceedings of the Eye Tracking Research and Applications Symposium. 2000. ACM Press.
161. Lau, J.S. and L. Huang, The prevalence effect is determined by past experience, not future prospects. Vision Res, 2010. **50**(15): p. 1469-74.
162. Wolfe, J.M. and M.J. Van Wert, Varying target prevalence reveals two dissociable decision criteria in visual search. Curr Biol, 2010. **20**(2): p. 121-4.
163. Pickhardt, P.J., et al., Assessment of volumetric growth rates of small colorectal polyps with CT colonography: a longitudinal study of natural history. Lancet Oncol, 2013. **14**(8): p. 711-20.

164. Boone, D.J., et al., CT colonography: external clinical validation of an algorithm for computer-assisted prone and supine registration. Radiology, 2013. **268**(3): p. 752-60.
165. Hampshire, T., et al., Endoluminal surface registration for CT colonography using haustral fold matching. Med Image Anal, 2013. **17**(8): p. 946-58.
166. Yushkevich, P.A., et al., User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage, 2006. **31**(3): p. 1116-28.
167. Li, P., et al., Registration of central paths and colonic polyps between supine and prone scans in computed tomography colonography: pilot study. Med Phys, 2004. **31**(10): p. 2912-23.
168. de Vries, A.H., et al., Feasibility of automated matching of supine and prone CT-colonography examinations. Br J Radiol, 2006. **79**(945): p. 740-4.
169. Summers, R.M., et al., Normalized distance along the colon centerline: a method for correlating polyp location on CT colonography and optical colonoscopy. AJR Am J Roentgenol, 2009. **193**(5): p. 1296-304.
170. Punwani, S., et al., Quantitative assessment of colonic movement between prone and supine patient positions during CT colonography. Br J Radiol, 2009. **82**(978): p. 475-81.
171. Plumb, A.A., et al., Increasing Navigation Speed at Endoluminal CT Colonography Reduces Colonic Visualization and Polyp Identification. Radiology, 2017. **284**(2): p. 413-422.
172. Wolfe, J.M., Use-inspired basic research in medical image perception. Cogn Res Princ Implic, 2016. **1**(1): p. 17.
173. Wood, G., et al., Exploring the impact of expertise, clinical history, and visual search on electrocardiogram interpretation. Med Decis Making, 2014. **34**(1): p. 75-83.
174. Giovinco, N.A., et al., A passing glance? Differences in eye tracking and gaze patterns between trainees and experts reading plain film bunion radiographs. J Foot Ankle Surg, 2015. **54**(3): p. 382-91.
175. Soh, B.P., et al., E-tutorial improves students' ability to detect lesions. Radiol Technol, 2013. **85**(1): p. 17-26.
176. Turgeon, D.P. and E.W. Lam, Influence of Experience and Training on Dental Students' Examination Performance Regarding Panoramic Images. J Dent Educ, 2016. **80**(2): p. 156-64.
177. van der Gijp, A., et al., How visual search relates to visual diagnostic performance: a narrative systematic review of eye-tracking research in radiology. Adv Health Sci Educ Theory Pract, 2017. **22**(3): p. 765-787.
178. Venjakob, A.C. and C.R. Mello-Thoms, Review of prospects and challenges of eye tracking in volumetric imaging. J Med Imaging (Bellingham), 2016. **3**(1): p. 011002.
179. Diaz, I., et al., Eye-tracking of nodule detection in lung CT volumetric data. Med Phys, 2015. **42**(6): p. 2925-32.
180. Rubin, G.D., et al., Characterizing search, recognition, and decision in the detection of lung nodules on CT scans: elucidation with eye tracking. Radiology, 2015. **274**(1): p. 276-86.
181. Kok, E.M., et al., Systematic viewing in radiology: seeing more, missing less? Adv Health Sci Educ Theory Pract, 2016. **21**(1): p. 189-205.

182. van Geel, K., et al., Teaching Systematic Viewing to Final-Year Medical Students Improves Systematicity but Not Coverage or Detection of Radiologic Abnormalities. J Am Coll Radiol, 2017. **14**(2): p. 235-241.
183. Plumb, A.A., et al., Small Polyps at Endoluminal CT Colonography Are Often Seen But Ignored by Radiologists. AJR Am J Roentgenol, 2015. **205**(4): p. W424-31.
184. Kelly, B.S., et al., The Development of Expertise in Radiology: In Chest Radiograph Interpretation, "Expert" Search Pattern May Predate "Expert" Levels of Diagnostic Accuracy for Pneumothorax Identification. Radiology, 2016. **280**(1): p. 252-60.
185. Venjakob, A.C., et al., Image Size Influences Visual Search and Perception of Hemorrhages When Reading Cranial CT: An Eye-Tracking Study. Hum Factors, 2016. **58**(3): p. 441-51.
186. Aizenman, A., et al., Comparing search patterns in digital breast tomosynthesis and full-field digital mammography: an eye tracking study. J Med Imaging (Bellingham), 2017. **4**(4): p. 045501.
187. Bertram, R., et al., Eye Movements of Radiologists Reflect Expertise in CT Study Interpretation: A Potential Tool to Measure Resident Development. Radiology, 2016. **281**(3): p. 805-815.
188. Kelly, B., et al., Influence of radiology expertise on the perception of nonmedical images. J Med Imaging (Bellingham), 2018. **5**(3): p. 031402.
189. Tourassi, G., et al., Investigating the link between radiologists' gaze, diagnostic decision, and image content. J Am Med Inform Assoc, 2013. **20**(6): p. 1067-75.
190. Gandomkar, Z., et al., iCAP: An Individualized Model Combining Gaze Parameters and Image-Based Features to Predict Radiologists' Decisions While Reading Mammograms. IEEE Trans Med Imaging, 2017. **36**(5): p. 1066-1075.
191. Drew, T. and L.H. Williams, Simple eye-movement feedback during visual search is not helpful. Cogn Res Princ Implic, 2017. **2**(1): p. 44.
192. Littlefair, S., et al., Does Expectation of Abnormality Affect the Search Pattern of Radiologists When Looking for Pulmonary Nodules? J Digit Imaging, 2017. **30**(1): p. 55-62.
193. Wen, G., et al., Comparative study of computational visual attention models on two-dimensional medical images. J Med Imaging (Bellingham), 2017. **4**(2): p. 025503.