COBIx PROTOCOL

COBIx: Multi-site validation study of the Colon and Rectal Endoscopic Biopsy (COBIx) reporting tool

ISRCTN Number:	TBC
Sponsor:	UNIVERSITY HOSPITALS COVENTRY AND WARWICKSHIRE NHS TRUST
Sponsor Number:	DS631023
Funding Body:	NIHR, Accelerated Access Collaborative
Funder Reference:	AI_AWARD02688
IRAS Number:	**330776**
Ethics Approval date:	06.11.2023 (REC) 20.11.2023 (HRA)
Version Number:	1.0
Date:	12th October 2023
Stage:	Final

This protocol has regard for current HRA guidance and content.

Protocol Version History:

Protocol Version Number	Protocol Version Date
V1.0	12th October 2023

This study is funded by the National Institute for Health Research (NIHR) Health Technology Assessment Programme [17/84/07]. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

CONTACT DETAILS

Sponsor	University Hospitals Coventry and Warwickshire NHS Trust University Hospital Coventry CV2 2DX Tel: 02476 966195 Email: ResearchSponsorship@uhcw.nhs.uk	**Chief Investigator**	**Professor David Snead** Consultant Pathologist Department of Pathology University Hospitals Coventry and Warwickshire NHS Trust University Hospital Coventry CV2 2DX Tel: 02476 968320 Email: David.Snead@uhcw.nhs.uk; David.Snead@pathlake.org
Research fellow	**Dr Harriet Evans** Research Fellow Department of Pathology University Hospitals Coventry and Warwickshire NHS Trust University Hospital Coventry CV2 2DX Email: Harriet.Evans4@nhs.net	**Project Manager**	**Rachel Flowers** Research Institutes Project Manager Research & Development University Hospitals Coventry and Warwickshire NHS Trust University Hospital Coventry CV2 2DX Tel: 024 7696 8650 Email: cobix@uhcw.nhs.uk
Principal Investigator	**Prof Nasir Rajpoot** TIA Centre Dept of Computer Sciences The University of Warwick Coventry CV4 7AL Email: N.M.Rajpoot@warwick.ac.uk	**Co-Investigator**	**Dr Fayyaz Minhas** TIA Centre Dept of Computer Sciences The University of Warwick Coventry CV4 7AL Email: Fayyaz.Minhas@warwick.ac.uk
Co-Investigator	**Dr Shan Raza** TIA Centre Dept of Computer Sciences The University of Warwick Coventry CV4 7AL Email: Shan.Raza@warwick.ac.uk	**TIA Centre Programme Manager**	**Johnathan Breddy** TIA Centre Dept of Computer Sciences The University of Warwick Coventry CV4 7AL Email: Johnathan.Breddy@warwick.ac.uk
Data Scientist	**Noorul Wahab** TIA Centre Dept of Computer Sciences The University of Warwick Coventry CV4 7AL Email: Noorul.Wahab@warwick.ac.uk	**WCTU Lead & Senior Trial Statistician**	**Prof Janet Dunn** Warwick Clinical Trials Unit The University of Warwick Gibbet Hill Road Coventry CV4 7AL Tel: 02476 575847 Email: J.A.Dunn@warwick.ac.uk

Trial Statistician	**Dr Louise Hiller** Warwick Clinical Trials Unit The University of Warwick Gibbet Hill Road Coventry CV4 7AL Tel: 02476 150179 Email: L.Hiller@warwick.ac.uk	**Trial Statistician**	**Dr Peter Kimani** Warwick Clinical Trials Unit The University of Warwick Gibbet Hill Road Coventry CV4 7AL Tel: 02476 151187 Email: Peter.K-U.Kimani@warwick.ac.uk
Trial Coordinator	**Donna Howe** Warwick Clinical Trials Unit The University of Warwick Gibbet Hill Road Coventry CV4 7AL Tel: 02476 150600 Email: D.L.Howe@warwick.ac.uk	**Senior Project Manager**	**Helen Higgins** Warwick Clinical Trials Unit The University of Warwick Gibbet Hill Road Coventry CV4 7AL Tel: 02476 151178 Email: H.Higgins@warwick.ac.uk
Health Economist	**Dr Henry Nwankwo** Assistant Professor in Health Economics Warwick Clinical Trials Unit Warwick Medical School University of Warwick Gibbet Hill Road Coventry CV4 7AL Tel: 02476 574872 Email: Henry.Nwankwo@warwick.ac.uk	**Principal Investigator**	**Dr Rebecca Brais** Cambridge University Hospitals NHS Foundation Trust Addenbrooke's Hospital Hills Road Cambridge CB2 0QQ Email: rebecca.brais@nhs.net
Principal Investigator	**Dr Paul Barrat** County Durham and Darlington NHS Foundation Trust Darlington Memorial Hospital Hollyhurst Road Darlington County Durham DL3 6HX Email: Paul.Barrett5@nhs.net	**Principal Investigator**	**Dr Prakash Konanahalli** Greater Glasgow Health Board Gartnavel Royal Hospital 1055 Great Western Road Glasgow G12 Email: prakash.Konanahalli@ggc.scot.nhs.uk
Principal Investigator	**Dr Kaushiki Dasgupta** North Tees and Hartlepool Hospitals NHS Foundation Trust University Hospital of North Tees Hardwick Road Stockton TS19 8BE Email: K.Dasgupta2@newcastle.ac.uk	**Principal Investigator**	**Dr Mohammad Ilyas** Nottingham University Hospitals NHS Trust City Hospital Hucknall Road NG5 1PB Email: Mohammad.Ilyas@nottingham.ac.uk

Principal Investigator	**Dr Chris Bacon** The Newcastle Upon Tyne Hospitals NHS Foundation Trust Freeman Hospital Freeman Road Hight Heaton Newcastle upon Tyne NE7 7DN Email: Chris.Bacon@newcastle.ac.uk	**Principal Investigator**	**Dr Richard Colling** Oxford University Hospitals NHS Foundation Trust John Radcliffe Hospital Headington Oxford OX3 9DU Email: Richard.Colling@ouh.nhs.uk
Principal Investigator	**Dr Kelvin Robson** The Royal Wolverhampton NHS Trust New Cross Hospital Wolverhampton WV10 0QP Email: Kelvin.Robson@nhs.net	**Principal Investigator**	**Dr Andrew Robinson** University Hospitals Coventry and Warwickshire NHS Trust University Hospital Coventry CV2 2DX Email: Andrew.Robinson@uhcw.nhs.uk
Principal Investigator	**Prof Ali Khurram** Sheffield Teaching Hospitals NHS Foundation Trust Northern General Hospital Herries Drive Sheffield S5 7AU Email: S.A.Khurram@sheffield.ac.uk	**Principal Investigator**	**Dr Vipul Foria** University Hospital Southampton NHS Foundation Trust Southampton General Hospital Tremona Road Southampton SO16 6YD Email: Vipul.Foria@uhs.nhs.uk
PPI Representative	**Sophia Turner** Email: sophiaturner@gmail.com		

TABLE OF CONTENTS

PAGE

LIST OF TABLES **PAGE**

LIST OF FIGURES **PAGE**

LIST OF ABBREVIATIONS

Abbreviation	Explanation
AI	Artificial Intelligence
AUROC	Area Under the Receiver Operating Characteristic
BSI	British Standards Institution
CI	Chief Investigator
CONSORT	*Con*solidated *S*tandards *o*f *R*eporting *T*rials
DCS	Department of Computer Science
EPR	Electronic Patient Record
GCP	Good Clinical Practice
H&E	Haematoxylin and Eosin
HRA	Health Research Authority
IRAS	Integrated Research Application System
iQC	Image Quality Control
ISRCTN	International Standard Randomised Controlled Trial Number
IVD-CE	In Vitro Diagnostic - Conformité Européene (European Conformity)
MHRA	Medicines and Healthcare products Regulatory Agency
MRC	Medical Research Council
NPV	Negative Predictive Value
PI	Principal Investigator
PPI	Patient & Public Involvement
PPV	Positive Predictive Value
REC	Research Ethics Committee
R&D	Research and Development
SMG	Study Management Group
SOP	Standard Operating Procedure
SSC	Study Steering Committee
TIA	Tissue Image Analytics
UHCW	University Hospital Coventry and Warwickshire
UK-CA	United Kingdom Conformity Assessed
UoW	University of Warwick
WCTU	Warwick Clinical Trials Unit
WSI	Whole-Slide Image

GLOSSARY

Term	Definition
Case	The whole case belonging to a patient, this may contain one or more specimens.
Specimen	A piece of tissue taken from specific part of the large bowel. The separate specimens taken from that patient form the case (e.g., often multiple biopsies are taken from different parts of the colon in a single case). The terms sample, specimen and biopsy all mean the same thing.
Slide	The specific slides produced from each specimen/specimens. Each specimen may be split into/ produce several slides (for example large specimens or if a pathologist has requested multiple levels), or multiple specimens may be placed on a single slide. The diagnosis of a specimen issued in the clinical report at each trust.
Reference diagnosis	The study pathologist's diagnosis at time of recruitment. This is considered the gold standard for algorithm comparison.
COBIx diagnosis	The diagnosis given by COBIx.
Ground truth diagnosis	The final diagnosis given to cases where there was discrepancy between the reference diagnosis and COBIx diagnosis, following consensus review of the case.

1. STUDY SUMMARY SCHEMA

Main study

Design: A multi-centre study comparing classification resulting from COBIx algorithm with pathologists' reporting of colon biopsy specimens.

Enrolment: 11,000 cases will be enrolled from 11 participating NHS sites (1,000 cases per site).

Primary objective: Comparing classification of normal/ abnormal categories.

Secondary objectives: Comparing classification of into neoplastic urgent/ neoplastic non-urgent/ non-neoplastic urgent/ non-neoplastic non-urgent categories.

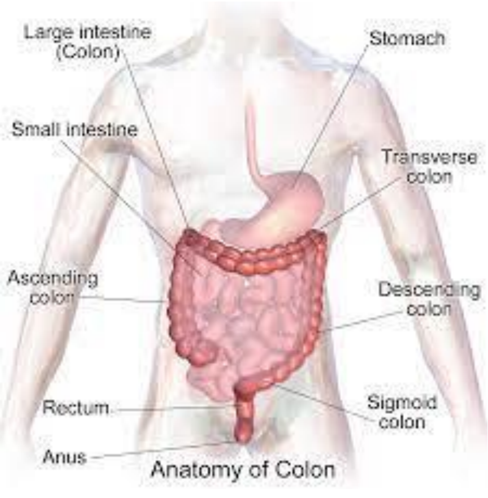
Data analysis: Validation of the COBIx algorithm by comparing COBIx and pathologists' classifications of the same specimens.

Sub study

Health economics: To assess the impact of COBIX algorithm on workflow productivity.

2. PLAIN ENGLISH SUMMARY

The colon (also called the large bowel) is part of our digestive system or gut. Sometimes diseases can form in the colon – for example colitis, Crohn's disease, and cancer. These can be serious and usually need to be treated.



Diseases are usually diagnosed by taking a sample of tissue (a biopsy) from the colon and having a pathologist (a doctor trained to examine tissue) look at it under a microscope. The pathologist can say whether a sample looks normal (healthy) or not.

Currently, around one third of these samples turn out to be normal. We are looking at ways of checking samples more efficiently whilst making sure we don't miss any which have disease and need treating.

New technologies using computer programmes (algorithms) are very promising. The one we are looking at in this study is called COBIx. In a previous study, the algorithm has been developed on thousands of colon samples and has been shown to be highly accurate.

We now need to test COBIx even more to make sure it is accurate wherever it is used, as every hospital has slightly different equipment and different ways of working.

For this study we plan to look at 11,000 colon cases from 11 different hospital trusts. For each one we will compare the COBIx result with the findings of a pathologist.

We want to find out whether COBIx is accurate and can be used safely in the future to allow patients to be diagnosed more quickly and safely.

3. BACKGROUND & RATIONALE

Histopathological diagnosis is a pivotal step in the diagnosis and treatment pathways of many major diseases. It provides the key diagnostic information and myriad characteristics of the diseased tissue important for deciding the future management plan. Advances, particularly around screening for early detection of cancer and improved life expectancy, are placing additional burden on already overstretched Cellular Pathology resources within the NHS. Early stage disease is more difficult to detect leading to more challenging and often increased numbers of biopsies and more data from these biopsies is needed to provide the best standard of care. In Cellular Pathology, 32% of consultants are over 55 and most of them expected to retire in 5 years; the number of new consultants is less than half the number expected to retire; 55% of the histopathology departments surveyed hold vacancies[1,2]. The result is escalating pressure on Cellular Pathology testing capacity (Figure 1), contributing to delays to diagnosis of cancer and other disease which worsens outcomes for patients[3]. Endoscopic biopsies from the large bowel account for 17.8% of the requests in NHS cellular pathology laboratories. Of these biopsies, 34.5% are reported as normal (Table 1). For normal slides, the pathologist contributes minimally to the care of the patient. Yet the scrutiny of the slides to deliver this verdict remains similar to slides containing disease, where the pathologist's expertise is crucial in guiding the clinical team on how to treat the patient.

Site	Histopathology requests	Large bowel biopsies (%)	Large bowel biopsies normal (%)
Coventry	41,771	4,877 (11.7)	1,680 (34.4)
Wolverhampton	52,008	9,708 (18.7)	4,140 (42.6)
Oxford	56,575	7,766 (13.7)	3,938 (50.7)
Nottingham	59,851	10,562 (17.6)	3,428 (32.4)
Newcastle	59,843	5,348 (8.9)	2,015 (37.7)
Durham	34,958	6,240 (17.8)	2,353 (37.7)
Glasgow	108,000	29,000 (13.9)	7,830 (27)
Total	413,006	73,501 (17.8)	25,384 (34.5)

Table 1 The numbers of colon and rectal biopsies performed in 2019 in comparison to overall workload
The proportion of biopsies reported as normal varies between sites ranging between 27 and 50.7%, the mean value is 34.5%.

Digital pathology systems[4,5]are recognised as important in addressing these problems and allow Artificial Intelligence (AI) tools to revolutionise delivery of Cellular Pathology[1,6-8]. Transition to digital pathology across the UK has been accelerated during the COVID-19 pandemic[9]. However, use of AI is limited and there is no existing AI addressing the screening of large bowel biopsies. The colon and rectal biopsy (COBI) tool is an exemplar project of the Industrial Strategy Challenge Fund centre of excellence PathLAKE[10], designed to efficiently and accurately screen normal colon biopsies in order to address this unmet need.

4. STUDY DESIGN

4.1 Aims and Objectives

The reporting and diagnosis of endoscopic large bowel biopsies in the United Kingdom is currently performed manually, mostly by consultant pathologists. This proposal aims to use AI pre-screening of slides to reduce workload and improve workflow. The colon and rectal biopsy reporting AI tool (COBIx) screens digitised whole-slide images (WSIs) of Haematoxylin and Eosin (H&E) stained biopsies sorting them into five groups: normal, abnormal neoplastic urgent, abnormal neoplastic non-urgent, abnormal non-neoplastic urgent and abnormal non-neoplastic non-urgent.

The proposal is to use this tool to remove normal slides from the pathologists' workload and to triage abnormal cases into urgent and non-urgent groups depending on the severity of disease detected. This will allow pathologists to focus on abnormal specimens, meaning patients with potentially serious disease are prioritised for immediate review. This should facilitate quicker diagnostic reporting, with greater priority given to serious disease, leading to faster treatment decisions. This helps deliver the aims of the NHS Long Term Plan[14] to diagnose cancer earlier and to support pathologists, clinicians, and patients by concentrating health resources on those patients that require treatment.

COBI has been developed as part of the PathLAKE project. The algorithm has been trained on ~3,400 WSIs and tested on 1,700 unseen slides delivering an area under receiver operating characteristic curve (AUROC) value of 0.96-0.99. These results demonstrate highly effective segregation of slides into three categories (normal, abnormal neoplastic, abnormal non-neoplastic), with high sensitivity and negative predictive values comparable to human pathologists. COBIx, the next iteration which includes individual cell recognition is now being introduced to deliver greater accuracy for the identification of non-neoplastic disease and triaging of cases into urgent and non-urgent categories. We have made careful provision for IVD-CE and UK-CA clearance, sharing with BSI and the MHRA our approach to the development of this tool. The multi-site study outlined in this protocol will provide the efficacy and safety data needed for regulatory approval, as well as key health economic data indicating the impact the technology will have in routine practice.

4.2 The COBIx Algorithm Version 1.0

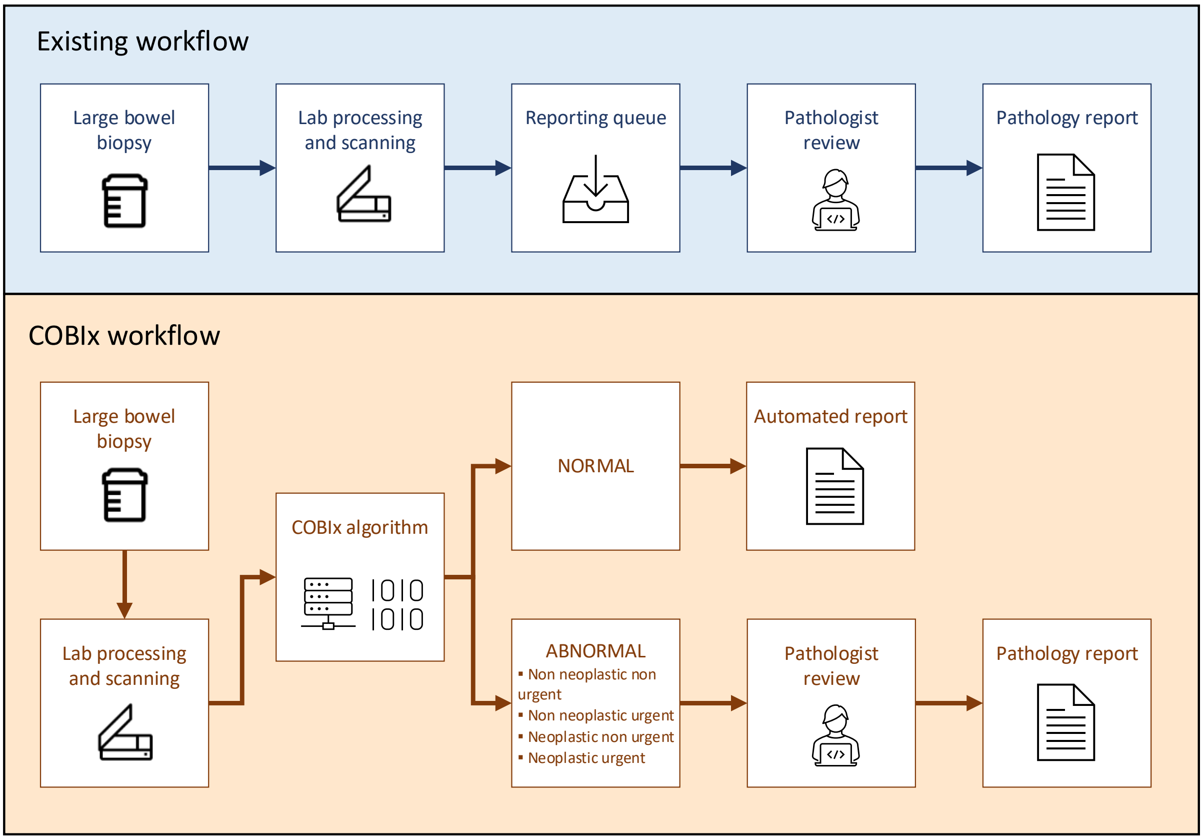


Figure 1 Overview of the COBIx algorithm for large bowel endoscopic biopsies.

Scanned WSIs of the biopsies are pre-screened by the COBIx algorithm which divides the biopsies into normal and abnormal outcome baskets. Abnormal biopsies are triaged into urgent and non-urgent groups depending on the pathology identified. Biopsies with a normal result trigger an automated report of normal tissue. These specimens would not be examined by the pathologist except at the request of the clinician or for quality assurance purposes.

The COBIx algorithm (Figure 1) is an in-vitro diagnostic medical device (computer software), intended for use on endoscopic large bowel biopsies from adult patients (over 18 years old) to screen normal from abnormal biopsies and assist in the triage of cases with serious disease for urgent pathologist review. Screening for normal is intended to provide a means of automated reporting of these specimens.

COBIx combines two parallel strands of AI technology: ***(a)*** Hover-Net, the state-of-the-art deep learning model providing individual cell recognition[11,12], and ***(b)*** iterative-draw-and-rank-sampling (IDARS) algorithm, a weakly supervised deep learning strategy recently published in The Lancet Digital Health[13], for the detection of colorectal carcinoma from routine H&E slides. Combining these algorithms predicts normal and the presence of neoplastic disease, invasive carcinoma and areas of acute inflammation and confidence levels for these predictions. The tool operates on digitised WSIs of colon biopsy slides stained with H&E.

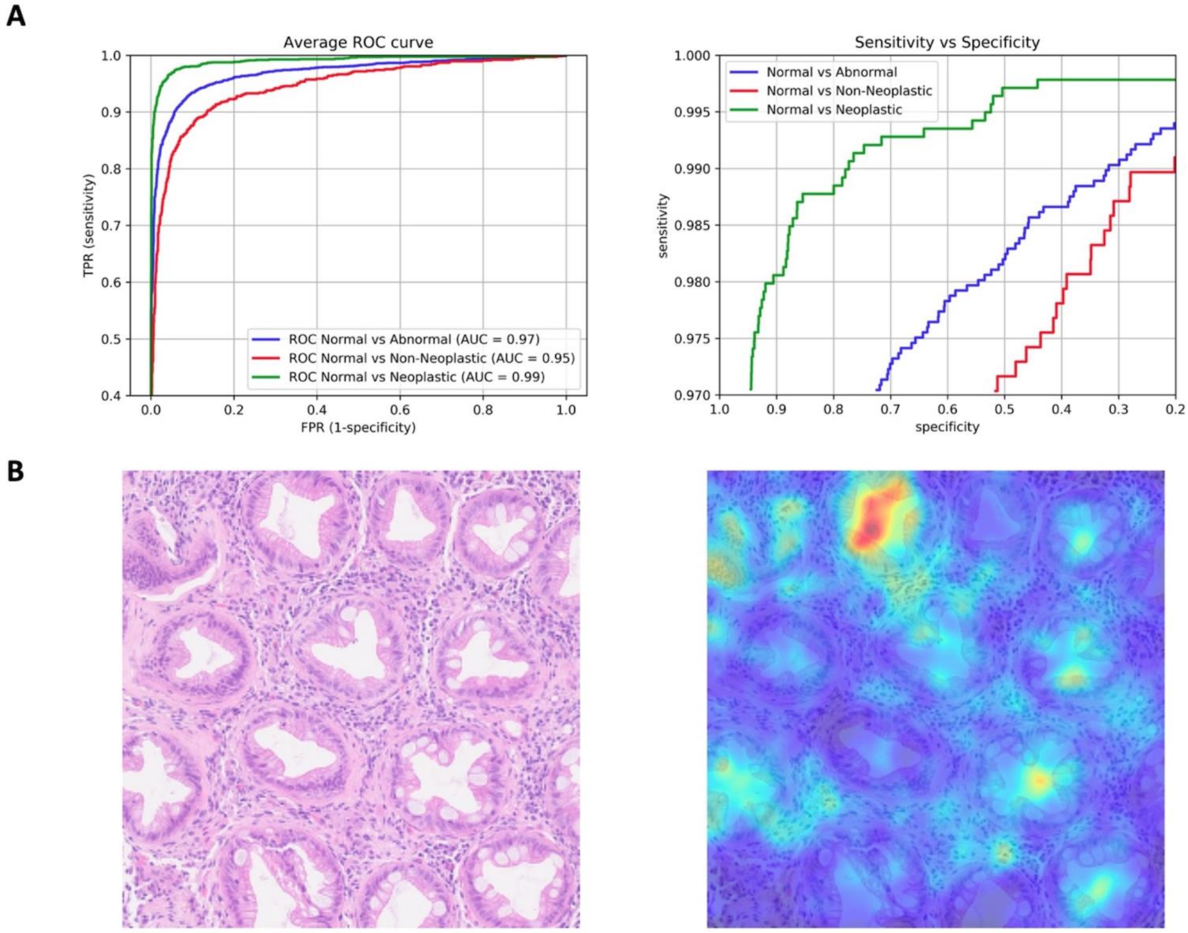


Figure 2 Results of the COBI algorithm based on IDARS a weakly supervised algorithm developed from slide deep learning of data with just the slide level diagnosis.

As shown in Figure 2A, there is excellent detection of neoplastic lesions by the algorithm, but detection of inflammatory conditions is less strong. Figure 2B shows the H&E image of a hyperplastic polyp with a corresponding heat map of the abnormality detected by the algorithm. (IDARS stands for iterative draw and rank sampling which is the approach taken to select informative patches in the training sets used in the algorithm development).

The PathLAKE project has created the prototype, on a dataset of over 8,000 WSIs of large bowel biopsies, with slide level diagnosis on 5,107 slides, divided into training and test sets (60% and 40% respectively). Using the IDARS algorithm alone at x5 magnification equivalent images, and hence only architectural morphological features, we have already achieved excellent separation of normal from neoplastic lesions and abnormal, both neoplastic and non-neoplastic lesions (see Figure 2).

Pathologist ground truth data, in triplicate on 150 selected slides has generated over 250,000 annotations that have been processed and added to the original model of Hover-Net[11] (see Figure 3). This is the largest dataset for nuclear instance segmentation in existence, with nearly half a million labelled nuclei in colon tissue[12] (Figure 3).

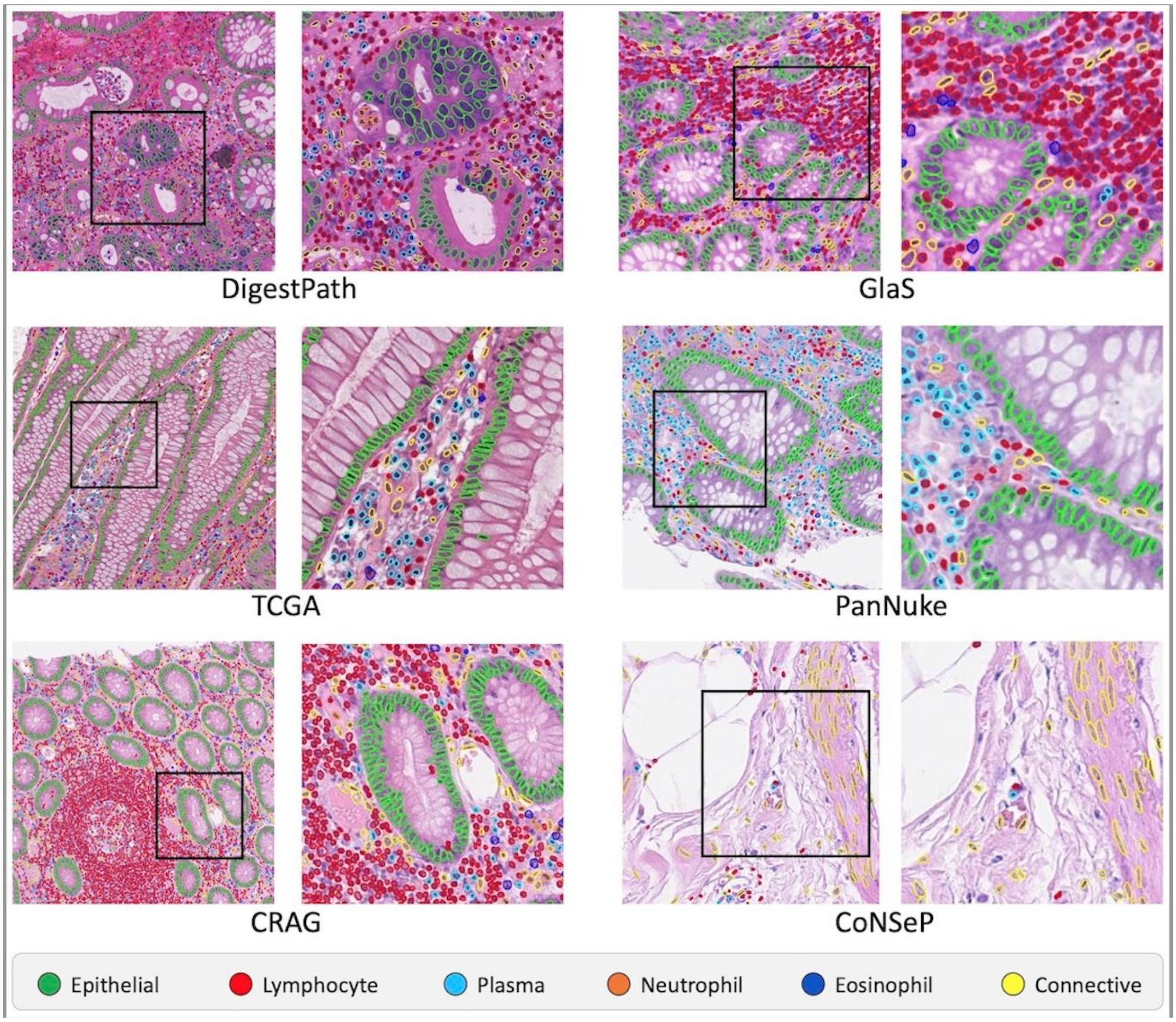


Figure 3 Selected fields of region and cell level annotation made by the Iguana algorithm which has been developed using pathologists' annotation of diagnostic regions and patched of cell level annotation in the training set of slides.

4.3 Medical device classification

COBIx is intended to recognise large bowel biopsies as normal or abnormal. It is intended that recognition of normal cases will generate an automated report. This report is designed to be posted onto the electronic patient record (EPR) and communicated to the patient, potentially using the NHS app on mobile devices. It is intended that abnormal cases will remain in the pathologist reporting workflow with annotations visible to the pathologist identifying the regions of interest (ROI) recorded as abnormal. COBIx will be classified as a class IIb device under Regulation (EU) 2017/745 Rule 11.

4.4 The Multi-site validation study outline

The study aims to recruit 11,000 patient cases from adults across 11 separate centres in the UK. All the H&E stained slides (from approximately 33,000 specimens) from these cases will be scanned (or in some cases, have already been scanned) to digital WSIs at the recruiting site using equipment used for routine diagnosis. Scanned images will be anonymised and transferred electronically to a storage housed within the TIA Centre. Once these digital slides have been transferred, they will be processed through the COBIx algorithm and classified into one of the three main categories and four subcategories. The results of the algorithm classification will be compared to the reference pathologist diagnosis.

4.5 Study size, case selection and case enrolment

The study will enrol 11,000 cases. Sample size for the analysis corresponds to the number of specimens and on average, it is expected that each case will have three specimens. In assessing sample size adequacy, we were conservative assuming that there will be one specimen per case, rather than the expected average of three per case. A sample size of 11,000 is adequate to estimate diagnostic measures (sensitivity, specificity, negative predictive value (NPV) and positive predictive value (PPV)) with high precision if the study will have a positive finding of obtaining sensitivity and NPV values of at least 0.98. Assuming we will observe NPV=0.98 and same number of normal cases by pathologists and COBIx algorithm, varying prevalence between 49% and 90% (observed data in seven sites in Table 1), the largest margin of error (1.96 x standard error) for any of the four diagnostic measures is 0.02 so that confidence intervals for all diagnostic measures will be very narrow. This sample size may seem large, but it is essential to deliver comprehensive assessment of the tool across all patient groups and the full range of pathological entities encountered in practice.

The study has been designed with the majority of cases (9,900) coming from retrospective cases, with 1,100 cases being recruited prospectively. Using retrospective cases is appropriate in this setting and ensures the volume of cases can be recruited and examined in the time available. It also ensures rare and unusual entities can be included so we can explore how these are handled by the algorithm. Prospectively recruited cases will be selected sequentially and will allow the study to collect health economics data, including the time taken for pathologists to examine normal biopsies and the broader impact on pathology workflow.

Selected cases are recruited and the specimen reviewed by the study pathologist for enrollment. The specimens are anonymised at the point of enrollment. The anonymisation process provides a unique case number, a part identifier, a block identifier and a slide identifying number. Where multiple specimens are combined into one block and all appear on one slide the process for anonymization and number is the same, but the slide receives multiple numbers to account for each specimen it contains. The fact that it is a multi-specimen slide is annotated in the study database. The orientation of the WSI will be used to segment the WSI into the component parts prior to the slide being processed through COBIx.

At the point of review, the study pathologist provides the diagnosis of each specimen in the case under the anonymized number on the study database. This is the **reference classification and diagnosis**. All cases are therefore enrolled with two pathologist reviews: the reporting pathologist and the study pathologist. These may be the same person. The timings for the health economics study are taken on the first review only.

To ensure representation from all ethnic groups, case selection will be divided between racial groups in the following proportions:

- White Caucasian groups 70%;
- Black, Asian and minority ethnic (BAME) groups, including racial group unknown 30%.

Prospective cases

Each site enrols one hundred sequential large bowel biopsy cases prospectively after the study starts.

Retrospective cases

Nine hundred cases are enrolled from the pathology archives at each site. Of these 100 cases is reserved for the categories of rare and unusual diagnoses. The remaining 800 cases are collected sequentially from a time period chosen by the enrolling site.

Classification of specimens

At enrolment, each specimen is classified by the recruiting pathologist into one of the categories shown below:

Diagnostic categories

1) **Normal** (any specimen which contains no evidence of disease)
2) **Non-neoplastic non-urgent** (any non-neoplastic abnormality which does not have acute inflammation - necrosis or ulceration)
3) **Non-neoplastic urgent** (any non-neoplastic abnormality which includes either acute inflammation or necrosis or ulceration)
4) **Neoplastic – non-urgent** (any neoplastic lesion which does not include either invasive carcinoma or other form of malignancy or high-grade dysplasia)
5) **Neoplastic urgent** (any neoplastic lesion which includes either high grade dysplasia or invasive carcinoma or other malignant diagnosis)

These categories match then the classification tree in figure 4.

All H&E slides from a case will be seen by the algorithm, but diagnostic categorisation will be done at the specimen level within a case.

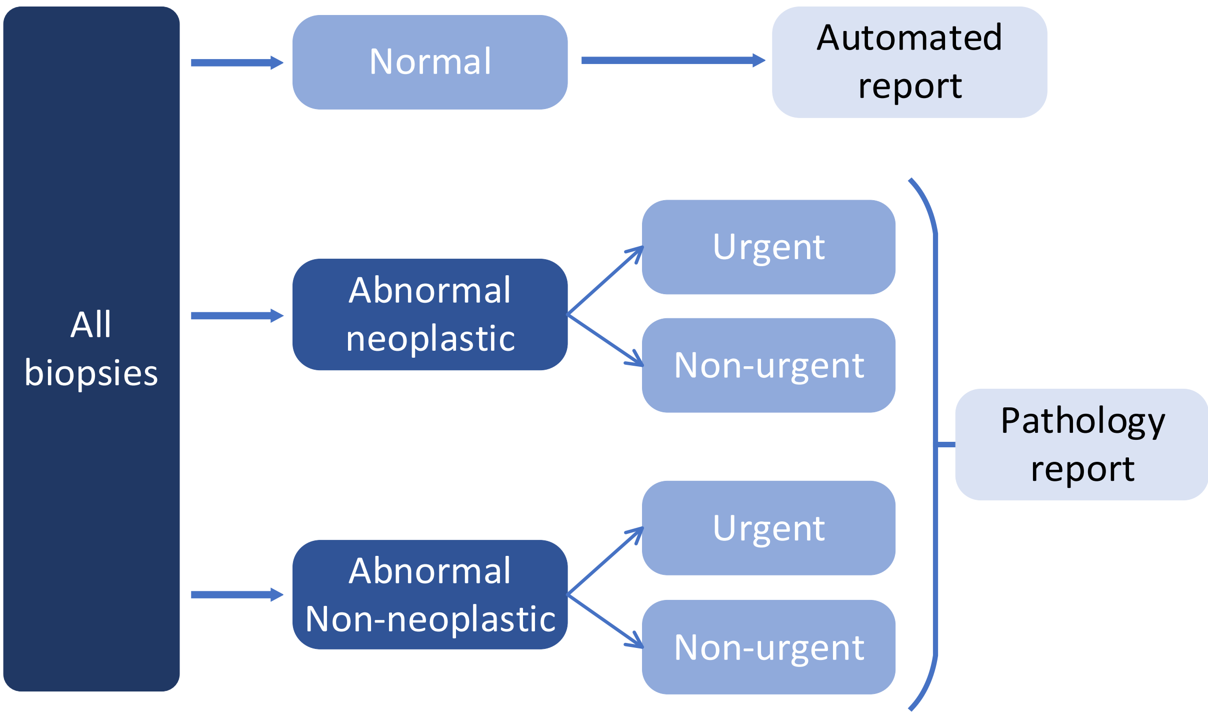


Figure 4 is a schematic diagram of the COBIx biopsy classification tree depicting how specimens will be classified in the study.

Both pathologists and COBIx classifies each specimen enrolled into one of the five diagnostic categories shown.

Case selection

Each site collects 1000 cases for the study in the following manner:
First, a search of the pathology archive for any of the diagnoses of interest (Cases of interest) listed below of any age and as many of these cases as can be found number X.
Second, a series of cases 900-X from any time period prior to August 2023 using an enrichment process to ensure 30% of cases come from minority ethnic groups, remaining cases are collected sequentially.
Third, a series of 100 cases collected during the trial period to be arranged with the coordinating centre.
Cases of interest, and other interesting and educational cases discovered during the course of the study, will be configured into a training set for the RCPath Portal.

For the cases enrolled into this study the site will review and ascertain the diagnosis of each specimen and categorise this into one of the 5 categories using the definitions provided.
This work is to be supervised by the principal investigator for the site, the work may be delegate to other staff (trainee pathologists, biomedical scientists, medical students etc. providing they have sufficient training and knowledge and are adequately supervised in this work) as appropriate.

Inclusion criteria

Large bowel endoscopic biopsies from adult patients (over 18 years)

Exclusion criteria

Specimens which are not from large bowel endoscopic biopsies
Incomplete cases with lost slides
Cases from person opting out via the NHS Data opt-out scheme
Faded or damaged slides
Incomplete or missing meta data

The following metadata is required for each case:

Site code
Laboratory number
Ethnic group
Gender
Age group
Procedure type
Tissue Type
Date of biopsy
Colonoscopy report
Clinical details
Specimen parts
Specimen blocks
Macroscopic report

Microscopic report
Diagnosis
SNOMED Code
Scanner type
Image resolution

Table 2 Metadata required for each case

Consensus review of discrepant samples

Where there is a discrepancy between the algorithm diagnostic category and pathologist diagnostic category the case will be reviewed by the consensus review team. A minimum of 3 pathologists will decide the correct or ground truth diagnostic category for each slide in the discrepant sample.

Diagnostic categories

1 **Normal** (any sample which contains no evidence of disease)
2 **Non-neoplastic – non-urgent** (any non-neoplastic abnormality which does not have acute inflammation - necrosis or ulceration)
3 **Non-neoplastic urgent** (any non-neoplastic abnormality which includes either acute inflammation or necrosis or ulceration)
4 **Neoplastic – non-urgent** (any neoplastic lesion which does not include either invasive carcinoma other form of malignancy or high-grade dysplasia)
5 **Neoplastic urgent** (any neoplastic lesion which includes either high grade dysplasia or invasive carcinoma or other malignant diagnosis).

Rare entities sought for enrichment:

Entity	Relevant SNOMED codes
Pseudolipomatosis	Not found
Pouchitis	Not found
Diversion colitis	Not found
Eosinophillic colitis	D5-41712, M-430401
Infectious aetiologies	-
Yersinia	L-1E400
Pseudomembranous C difficile colitis	L-14137, M-417801
Protozoan infections (amoebic colitis)	DE-50090, L-50001, L-515001
Histoplasmosis	DE-40700
Cryptosporidium	L-52400
Cytomegalovirus	L-36500, DE-32610
Adenovirus	L-35800
Mycobacterium avium complex	L-21815
Mycobacterium tuberculosis	L-21907
HIV associated giant cells	L-352101
Malakoplakia	M-44170, D5-44360, D5-45380
Parasites (e.g. T trichiura, E vermicularis, schistosomiasis, strongloidiasis)	L-50000, L-550001, L-56600, L-56601, L-55300, L-58120
Ischaemic colitis	F-39340, D5-41210
Vasculitis	D3-80650

Amyloidosis	M-97691,M-551002, D6-94500
Mastocytosis	M-97401
Langerhans cell histiocytosis	M-97513, D6-64952, M-97511, M-77870
Radiation colitis	M-11600, DD-64100, M-11620
Graph versus host disease	DC-50050
Diverticular disease associated colitis	M-40630, D5-43220, D5-43210, D5-43204
Mucosal prolapse conditions	-
Solitary rectal ulcer syndrome	Not found (D5-45420 = ulcer of rectum)
Inflammatory cloacogenic polyp	Not found
Diverticular disease associated polyps (polypoid prolapsing mucosal folds)	Not found
Cap polyposis	Not found
Juvenile polyps (sporadic and syndromic)	M-75662
Peutz-Jeghers polyps	M-75660
Endometriosis	M-76500
Colonic xanthoma	M-553002
Gastric heterotopia	M-00370, M-260004 (both non specific heterotopia)
Pulse granuloma	Not found (M-44002 - granuloma NOS)
Mucosal pneumatosis/ Pneumatosis cystoides intestinalis	D5-43320, M-333801
Leiomyoma	M-88900
Neural tumours including	-
Ganglioneuroma	M-94900
Neurofibroma	M-95400
Benign fibroblastic polyps/ perineuroma	M-95710
Mucosal schwann cell hamartoma	Not found
Benign epithelioid nerve sheath tumours/ epithelioid schwannoma	M-95600, M-956005
Neuroma	M-95700
Granular cell tumour	M-95800
Lipoma	M-88500
Vascular tumours including	-
Haemangioma	M-91200
Lymphangioma	M-91700
Kaposi sarcoma	M-91403, D3-F0843
Angiosarcoma	M-91203, M-912031
Gastrointestinal stromal tumour	M-89363, M-89361
Signet ring cell carcinoma	M-62140, M-849031, M-84903, M-84906
Metastatic malignancy NOS	M-80006
Metastatic adenocarcinoma	M-81406
Metastatic melanoma	M-87206
Metastatic sarcoma	M-88003
Metastatic SCC	M-80706
Lymphoma	M-95903, M-96993, M-95913, M-97023
Neuroendocrine tumours (of all types)	M-82403, M-80413, M-80133

Table 3 Rare entities sought for enrichment

4.6 COBIx algorithm analysis of the study slides

Once the cases have been anonymized, the WSIs and clinical metadata are transferred to the COBIx study section of the Tissue Image Analytics (TIA) Centre. The slides are then queued for processing through the COBIx algorithm. The COBIx classification of the specimens is added to the database and an automated lookup table compares the results to the reference diagnosis. Discrepant results are queued for pathologist re-review and ground truth diagnosis. TIA Centre will be working with Histofy Ltd for generating the predictions for the model. The TIA Centre will also be able to request non-diagnostic technical metadata for the cases received for the study from Warwick CTU as needed.

Before the slides are processed by COBIx algorithm, an automated image quality control (iQC) will be run on all the slides. The main purpose of this iQC check will be to identify images of insufficient quality to be processed by COBIx and would need to be rescanned or replaced. The checks will include detection of sufficient tissue area, blur, out of focus, tissue folding, and stain issues.

4.7 Discrepant specimens and the Ground Truth

Where there is a discrepancy between the algorithm diagnostic category and pathologist diagnostic category the case will be reviewed by the consensus review team. Discrepancies will be identified by WCTU (see Figure 5 below) by comparing the enrollment data from sites and the COBIx classifications of each specimen provided by the TIA Centre and/or Histofy. The WCTU will then forward the details onto the study Research Fellow and consensus group for review.

A minimum of 3 pathologists will decide the correct or ground truth diagnostic category for each slide in the discrepant specimen. If this should require additional laboratory work either in the form of deeper sections, special stains or immunocytochemistry this is requested by from the submitting site.

The consensus view of the study pathologists is the ground truth diagnosis. If this differs from the reference diagnosis, then the PI of the submitting site is informed of the difference, so that this can be brought to the attention of the clinical team. They will then consider if the difference has any impact on management of the patient and take whatever action they deem appropriate.

4.8 Specimen pathway

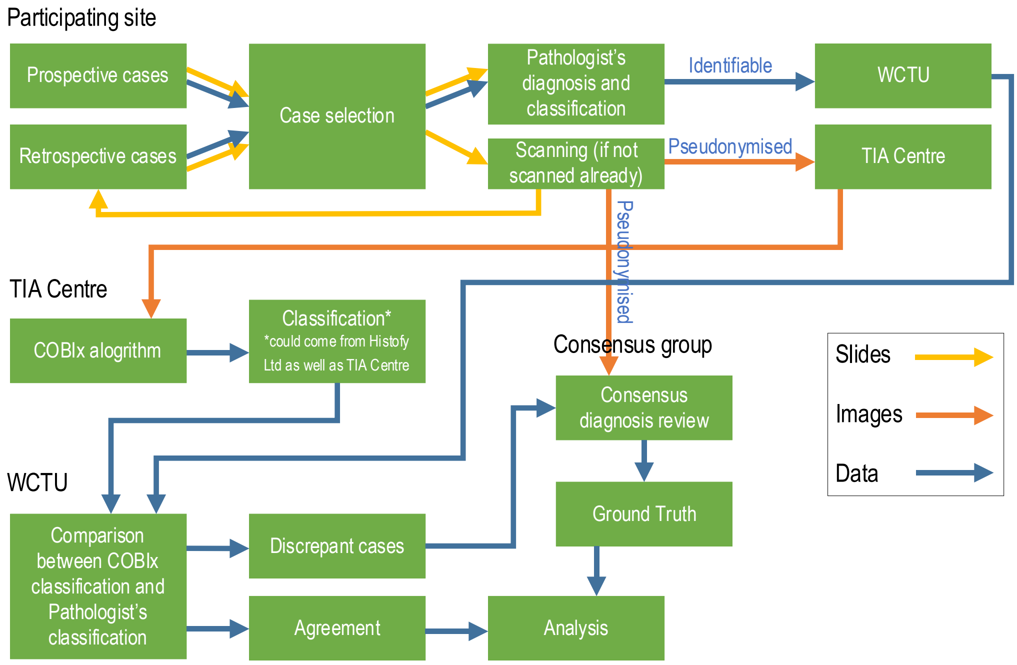


Figure 5 Specimen pathway

4.9 Outcome measures

The study measure is categorisation of diagnosis of a specimen into either normal or abnormal and, secondarily for abnormal specimens, into one of the four abnormal categories listed above. Reference diagnosis category by site pathologists (gold standard) will be compared to COBIx algorithm diagnosis category (experimental/intervention) using the following study outcomes.

Primary outcomes: For the validation part of the study, using a pre-specified COBIx algorithm threshold, the outcomes will be sensitivity, specificity, negative predictive value (NPV) and positive predictive value (PPV) to detect normal versus abnormal (all categories other than normal including changes of uncertain significance). For all possible COBIx algorithm thresholds, the outcomes will be receiver operating characteristic (ROC) and precision-recall (PR) curves.

Secondary outcomes: Agreement on abnormal categories, neoplastic versus non- neoplastic.

Tertiary outcomes: Agreement on urgent versus non-urgent categories (invasive carcinoma and or high-grade dysplasia versus non-invasive neoplasia, and acute inflammation, necrosis, ulceration versus all other pathology).

Additionally, we will use items recorded in clinical details and /or endoscopic findings to reduce the possibility and impact of false negative results from the COBIx algorithm and to development the safest possible reporting pathway in practice.

5. TRAINING

Site initiation

Each site will have a site initiation where the Principal Investigator and their team participate in an induction session. This was carried out over video conference.

A checklist will be completed for all sites to confirm that pre-activation activities are completed. Support is offered by relevant study team members to staff at participating sites to ensure they remain fully aware of study procedures and requirements. Working instructions will be provided for technical staff and pathologists. Additional support and training is offered to sites where necessary.

6. STATISTICAL ANALYSIS

The aim of the study is to validate the COBIx diagnostic tool that will be developed before the study begins. We will compute diagnostic measures of sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV). The gold standard in the analysis is a pathologist's reference diagnosis. A positive result will be the ability of the COBIx diagnostic tool to screen out normal cases. For patient safety, the desire is for the tool to achieve this whilst minimising false negatives and so very high sensitivity and very high NPV are desired. NPV is a function of prevalence of abnormal cases. The prevalence in the study will not be representative because there will be enriching for rare cases (e.g., by ethnicity). However, very high NPV is desired for any case mix. COBIx will be considered a good tool if sensitivity and NPV are above 0.98. Specificity will quantify the proportion of normal cases that are screened out and hence the saving in pathologists' workload.

For the cases that are abnormal on both the COBIx algorithm and reference diagnosis, based on the categorisation "neoplastic" and "non-neoplastic", we will compute kappa statistic. We will use the kappa statistic to assess agreement between COBIx and pathologist diagnosis in categorising abnormal cases as neoplastic or non-neoplastic. Also, taking the gold standard as a pathologist's categorisation of an abnormal case as neoplastic or non-neoplastic, we will compute sensitivity, specificity, and other diagnostic measures.

For the cases that are abnormal on both the COBIx algorithm and reference diagnosis, based on the categorisation "urgent" and "non-urgent", we will compute kappa statistic. We will use this kappa statistic to assess agreement between COBIx and pathologist diagnosis in categorising an abnormal case as urgent or non-urgent. Also, taking a pathologist's reference diagnoses as the gold standard, we will compute sensitivity, specificity, and other diagnostic measures. Comparison of categorisation as urgent and non-urgent will be based on all abnormal cases and in subgroup analyses of separately analysing neoplastic cases and non-neoplastic cases.

Subgroup analysis will repeat the above set of analyses for clinically important patient subgroups such as categories based on sex, ethnicity, age and site of the biopsy.

The validation part of the study that corresponds to the analysis described above will be based on a pre-specified COBIx algorithm threshold obtained using external data. Additional analysis will consider diagnostic ability of all COBIx algorithm predictive scores and so we will perform ROC curve analysis and PR curve analysis.

The performance of the predictive algorithm will be determined by Warwick Clinical Trials Unit (WCTU) using the reference diagnosis received from different sites and the class-wise prediction scores generated by different variants of the predictive algorithm.

An interim analysis is planned around month 18. The study will stop for futility if sensitivity/NPV is below a pre-specified value. The exact time for the interim analysis and the exact value of futility value will be pre-determined by the Study Steering Committee (SSC). The final analysis will adjust for the possibility of stopping for futility at interim analysis. A detailed and comprehensive statistical analysis plan (SAP), that will include how the final analysis will adjust for the possibility of stopping for futility, will be written and approved before performing the interim analysis.

7. HEALTH ECONOMICS EVALUATION

For prospective cases, submitting sites will record the length of time pathologists spend examining the H&E sections of the case at the first sitting. The time taken to examine each slide is recorded. For normal cases, the total time taken to examine all the slides, complete the report and authorise the case is recorded. If the pathologist has secretarial help to report the case, then the time taken to dictate the report and then to correct and authorise the report are recorded separately and then added to create one total time. For cases where some but not all the slides are normal, the time taken to examine normal slide(s) is recorded separately. We will undertake a micro-costing exercise to determine resources required to generate a normal report by pathologists and COBIx . We will explore the effect of COBIx on pathology workflow, including the impact of triage of specimens into diagnostic categories on patients referred to endoscopic biopsy. We will further explore the impact of COBIx on long term costs and outcomes on patients eligible for colon and endoscopic screening for colorectal cancer.

8. DATA MANAGEMENT

8.1 Data collection and data storage

At WCTU

Data will be collected at sites using a web application (database) for access and storage by WCTU. The web application will be developed by the Programming Team at WCTU and all specifications (i.e. database variables, validation checks, screens) will be agreed between the programmer and research team.

At TIA Centre

Pseudonymised WSI files will be transferred from the Partner Trusts to the TIA Centre using a Secure File Transfer Protocol (SFTP), such as FileZilla. The exception to this will be for UHCW which will use the secure link to the Medical School enabling access to the University network.

Once received, the files will be transferred to University of Warwick (UoW) IT Services data shares located in one data centre. A copy of the data will also be held in a separate IT Services data centre as a backup.

8.2 Confidentiality

At WCTU

All essential documentation and study records will be stored by WCTU in conformance with the applicable regulatory requirements. Access to stored information will be restricted to authorised personnel. An audit may be arranged at a site if the Study Management Group feels it is appropriate. Audits will be conducted by an independent team, determined by the Study Management Group.

8.3 Data shared with third parties

Data will be made available to third parties on approval.

8.4 Essential Documentation

A Study Master File will be set up according to WCTU SOP 11 (Essential Documentation: Creation and maintenance of Trial Master and Investigator Site Fles) and held securely at WCTU.

WCTU will provide the necessary documents to all participating centres involved in the study.

8.5 Archiving

At WCTU

Anonymised data will be held for a period of 10 years after completion of the study. Access to the study documentation will be restricted to named individuals within the study team with express permission from the Chief Investigator.

At TIA Centre

Whole slide images and study data will be archived for 10 years after completion of the study so that they can be access for future studies.

At Histofy

Histofy will be responsible for archiving additional data needed to support review by regulatory bodies. Archived data will be stored for 10 years.

9. INFORMED CONSENT

Patient cases will be recruited without consent and anonymised at the point of enrolment. The NHS data opt-out will be checked at point of enrolment by enrolling site, and those patients that have opted-out will not have their data included in the study.

10. PATIENT AND PUBLIC INVOLVEMENT (PPI)

In the design of the study, we discussed with PPI members , the practicalities around using prospective and retrospective cases, and the ethical considerations around the lack of patients' consent. They reviewed the feedback given on the stage 1 application and contributed to the changes made in the submission for stage 2.

We have encouraged engagement of our PPI colleagues in this study by inviting them to sit on the SSC and we will be providing the training necessary to enable them to carry out this task. We have consulted widely with interested patients and lay members of the public on the lay summary relevant to this study.

PPI members have also been involved in examining the justification of the study costs, and reviewing where efficiencies have been made and the value for money offered by this study.

PPI members will be invited to take part in the working group sessions to be held jointly between clinicians and PPIE members using role play techniques. In this work package we will examine the concept of AI being used to assess patient samples and how patients might react to this, as well as the potential impact of automated reporting of large bowel endoscopic biopsies on patients and the clinical teams caring for them. This work will be delivered by focus groups of PPIE members and clinicians meeting to explore scenarios developed by the Pathologists and computer scientists which are modelled on real examples of algorithm results. This work will be led by the Warwick Clinical Trials Unit. The purpose of this work being to establish how patients interact with clinicians if they receive results without follow-up appointments scheduled. What kind of additional support might be needed in order to ensure patients are fully aware of what should happen next, and how this support might need to be altered to accommodate for different patient groups, defined for example by age, gender, ethnicity, educational needs etc. PPIE members will also be involved in role play with clinicians to discuss the problem of false negative diagnoses, and how the risk of this is being managed and how it compares with the risks already present in the system.

11. STUDY ORGANISATION AND OVERSIGHT

11.1 Sponsor and governance arrangements

UHCW has agreed to act as sponsor for this trial and will undertake the responsibilities of sponsor as defined by the UK Policy Framework for Health and Social Care Research and ICH Good Clinical Practice. An authorised representative of the Sponsor has approved the final version of this protocol with respect to the trial design, conduct, data analysis and interpretation and plans for publication and dissemination of results. As sponsor, UHCW provides indemnity for this trial and, as such, will be responsible for claims for any negligent harm suffered by anyone as a result of participating in this trial. The indemnity is renewed on an annual basis and will continue for the duration of this trial.

11.2 Ethics and Health Research Authority Approval

The study will be conducted in accordance with all relevant regulations.

Health Research Authority advice (and approval, if needed) will be sought prior to the study commencing. Digital pathology algorithm development (REC Ref 5/NW/0843, IRAS project ID: 189095) has been used to develop the original COBI algorithm. The main ethical issue associated with this study is the use of patient data in anonymised form without consent except for following the NHS data opt-out. This is in line with existing guidelines and follows current practice. It would be impractical to gain consent for the numbers of cases included in the study given the majority of cases are archived. The use of anonymised data presents no risk to the patients involved.

11.3 Responsibilities

Principal Investigator (PI):

Each site needs to provide a pathologist to act as the site Principal Investigator. This person takes responsibility for the following tasks:

- Overseeing the recruitment of all case needed
- Ensuring each case is checked and reviewed at enrolment to confirm that the pathology present matches the diagnostic category
- Coordinating the delivery of the metadata and diagnostic category data for the cases recruited
- Overseeing the export of WSIs for the algorithm to compute.
- Providing a GI pathologist to take part in the consensus review of discrepant specimens

Chief Investigator (CI):

The Chief Investigator's responsibilities include, but are not limited to:

- Ensuring that the study is conducted as set out in the protocol and supporting documents
- Delegating study related responsibilities only to suitably trained and qualified personnel and ensuring that those with delegated responsibilities fully understand and agree to the duties being delegated to them
- Allowing access to source data for monitoring, audit, and inspection
- Ensuring the study is conducted in accordance with GCP principles

Study Management Group (SMG):

The Study Management Group, consisting of the study staff and pathologists involved in the day-to-day running of the study, will meet regularly throughout the project. Significant issues arising from management meetings will be referred to the Study Steering Committee or Principal Investigators, as appropriate.

The SMG's responsibilities include, but are not limited to:

- Coordinating development of protocol and study management documents
- Correspondence with study funder (NIHR-HTA)
- Setting up and maintaining the Study Master File
- Ensuring necessary approvals are in place before the start of the study
- Providing training to study personnel
- Providing data management support
- Producing study progress reports and coordinating SSC meetings and minutes
- Ensuring data security and quality and ensuring data protection laws are adhered to
- Ensuring complete records are in place for audit and monitoring purposes
- Ensuring the study is conducted in accordance with GCP guidelines
- Archiving all original study documents in line with UHCW NHS Trust policy

The full remit and responsibilities of the SMG will be documented in a Charter which will be signed by all members.

Study Steering Committee (SSC):

The study will be guided by a group of respected and experienced personnel and researchers as well as at least one 'lay' representative. The SSC will have an independent Chairperson. Face to face meetings will be held at regular intervals determined by need but not less than once a year. Routine business is conducted by email or teleconferencing. The SSC will take responsibility for:

- Major decisions such as a need to change the protocol for any reason
- Monitoring and supervising the progress of the study
- Reviewing relevant information from other sources
- Informing and advising on all aspects of the study

The full remit and responsibilities of the SSC will be documented in the Committee Charter which will be signed by all members.

Funder:

The study is funded by the National Institute for Health Research, Accelerated Access Collaborative. The design and management of the study are independent of the funder, however regular updates will be forwarded in study 'Progress report task' within the NETSCC Management Information System (MIS) portal.

11.4 Study Management

Each site appoints a pathologist to be the principal investigator. The chief investigator, PIs and Project Manager form the study management group. Once the study has been started, the study management group will meet regularly to review progress of the study.

The Sponsor and CI will finalise the study protocol. The WCTU, TIA Centre and PIs will review the study protocol and arrangements for how the study is to be conducted, record keeping and record and investigate any study violations. The WCTU will arrange site initiation with each site prior the study starting.

11.5 Monitoring, audit and inspection

The study is constructed around 11 collaborating centres each of which will select the study cases, with an additional further site reporting. The study will be managed by the Study Steering Committee (SSC), which has an independent chair. The data collection and storage are managed by WCTU and TIA.

Statistical analysis is being performed by Dr P Kimani and this is monitored by WCTU. An interim assessment of the project shall also be useful in reducing the risk. In case of recruitment issues, WCTU will be supported by the data science team in TIA Centre.

A Study Monitoring Plan will be developed and agreed by the Study Management Group (SMG) and SSC based on the study risk assessment which may include on site monitoring.

11.6 Indemnity

NHS indemnity covers NHS staff, medical academic staff with honorary contracts, and those conducting the study. NHS bodies carry this risk themselves or spread it through the Clinical Negligence Scheme for Trusts, which provides unlimited cover for this risk. The UoW provides indemnity for the design of the research protocol and conduct of the study.

11.7 Study timetable and milestones

• Finalisation of study protocol • Gain HRA approval • Site recruitment and contracts	Sep 2023 Month 6
• Set-up and opening sites • Data collection and slide scanning commenced	Dec 2023 Month 9
• Oversight committee chosen and nominations sent to NIHR	Feb 2024
• First SSC meeting • All sites open	March 2024 Month 12
• Retrospective cohort recruitment 50% complete • Prospective cohort recruitment commences • Interim analysis of study data	September 2024 Month 18
• Retrospective cohort 75% complete • Prospective cohort recruitment ongoing • Interim analysis of COBIx algorithm in cohort study	December 2024 Month 21
• Retrospective cohort recruitment complete and review of discrepant cases. • Prospective cohort 40% complete • SSC meeting	March 2025 Month 24
• Collection of prospective cohort cases (100% complete) • Review discrepant cases prospective cohort • Clinical study results • Health economics study results	September 2025 Month 30
• Study end date	31 March 2026 Month 36

12. END OF STUDY

The completion of the resolution of discrepant cases will be considered as end of study.

The study will be stopped prematurely if funding for the study ceases.

The Health Research Authority will be notified in writing within 90 days when the study has been concluded or within 15 days if terminated early.

Table 4 Study timetable and milestones

13. DISSEMINATION AND PUBLICATION

13.1 Dissemination

All data arising from the conduct of this study will remain the property of University Hospitals Coventry and Warwickshire NHS Trust. All efforts will be made to ensure that the results arising from the study are published in a timely fashion, in established peer-reviewed journals. Results will be disseminated via internal and external conferences and seminars, newsletters, and via interested groups, including local healthcare commissioning groups.

The results of the study will be reported first to study collaborators. The main report will be drafted by the study co-ordinating team, and the final version will be agreed by the Study Steering Committee before submission for publication, on behalf of the collaboration.

The success of the study depends on the collaboration of pathologist, technicians and researchers from across the UK. Equal credit will be given to those who have wholeheartedly collaborated in the study.

The study will be reported in accordance with the Consolidated Standards of Reporting Trials (CONSORT) statement (www.consort-statement.org).

13.2 Publication

It is planned to publish the findings in peer review, open access publications, following presentation at national and international conferences.

In addition to these publications, UHCW, the lead trust and sponsor, plans a series of press releases and media interviews on the progress and findings of the study.

We will develop a strategy in consultation with our PPI group and UHCW/UoW Communications team (e.g. a lay summary of the findings available on the study websites, and dissemination through social media) to help patients, patient representatives and wider public learn about the project's findings.

14. INTELLECTUAL PROPERTY

Intellectual property created in this project is non-severable from the existing intellectual property which is held jointly by the UoW and UHCW NHS Trust. Agreement for the onward management of this IP are covered in the project consortium agreement and with funding contract with NIHR, who are funding this study.

15. REFERENCES

1 Pathologists, R. C. o. Meeting Pathology Demand - Histopathology workforce census., (Royal College of Pathologists, 2018).

2 Bainbridge, S., Cake, R., Meredith, M., Furness, P. & Gordon, B. Testing times to come an evaluation of pathology capcity across the UK., (Cancer Research UK, 2016).

3 Hanna, T. P. *et al.* Mortality due to cancer treatment delay: systematic review and meta-analysis. *BMJ* **371**, m4087, doi:10.1136/bmj.m4087 (2020).

4 Mukhopadhyay, S. *et al.* Whole Slide Imaging Versus Microscopy for Primary Diagnosis in Surgical Pathology: A Multicenter Blinded Randomized Noninferiority Study of 1992 Cases (Pivotal Study). *Am J Surg Pathol* **42**, 39-52, doi:10.1097/PAS.0000000000000948 (2017).

5 Snead, D. R. *et al.* Validation of digital pathology imaging for primary histopathological diagnosis. *Histopathology* **68**, 1063-1072, doi:10.1111/his.12879 (2016).

6 Amess, D. (http://www.appghep.org.uk/meetings-2/, http://www.appghep.org.uk/meetings-2/, 2021).

7 Salto-Tellez, M., Maxwell, P. & Hamilton, P. Artificial intelligence-the third revolution in pathology. *Histopathology* **74**, 372-376, doi:10.1111/his.13760 (2019).

8 Cui, M. & Zhang, D. Y. Artificial intelligence and computational pathology. *Lab Invest* **101**, 412-422, doi:10.1038/s41374-020-00514-0 (2021).

9 Browning, L. *et al.* Digital pathology and artificial intelligence will be key to supporting clinical and academic cellular pathology through COVID-19 and future crises: the PathLAKE consortium perspective. *J Clin Pathol*, doi:10.1136/jclinpath-2020-206854 (2020).

10 Department for Business, E. a. I. S. (https://www.gov.uk, Gov.UK, 2018).

11 Graham, S. *et al.* Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Med Image Anal* **58**, 101563, doi:10.1016/j.media.2019.101563 (2019).

12 Graham, S. *et al.* Lizard: A Large-Scale Dataset for Colonic Nuclear Instance Segmentation and Classification. *Pre-print* (2021).

13 Bilal, M. *et al.* Novel deep learning algorithm predicts the status of molecular pathways and key mutations in colorectal cancer from routine histology images. *medRxiv*, 2021.2001.2019.21250122, doi:10.1101/2021.01.19.21250122 (2021).

14 NHS. The NHS Long Term Plan. (NHS, 2019).

16. APPENDICES

16.1 Companies involved in the development of COBIx

The **TIA Centre** at the UoW is a university research centre based in the Department of Computer Science (DCS). It was established in Jan 2021 to tackle some of the major challenges in systematic data-driven mining of an increasing deluge of cancer image data as well as the associated clinical and genomic data. Our ultimate aim is to develop cutting-edge AI technologies to assist pathologists in diagnosing cancer more efficiently and empower the oncologists with reliable information to select optimal personalised treatment for cancer patients.

Current research in the TIA Centre is focussed on the application of image analysis and machine learning algorithms in order to further our understanding of the biology and entangled histological patterns of complex diseases such as cancer. We strive to be a hub of research excellence in the area of computational pathology and associated research areas, in order to tackle grand challenges in cancer diagnostics and treatment with a multi-disciplinary team of researchers and to make positive impact on the lives of cancer patients. Our research thrives on a growing network of collaborations with the academia, NHS hospitals and industry.

Histofy is a spin-out from the UoW, specifically from the TIA Centre that specialises in the development of image analysis solutions for pathology. With over 250 publications in leading journals and conferences from our research spanning over 20 years, we thoroughly understand the process of technology development from conception to validation, placing us in a prime position to further develop and exploit AI technologies and integrate them in clinical workflows.