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

STATISTICAL ANALYSIS PLAN

ISRCTN:	ISRCTN96000018
Funding Body:	NIHR, Accelerated Access Collaborative
Ethics Approval:	06.11.2023 (REC) 20.11.2023 (HRA)
SAP Version:	1.0
Date:	21-04-2026
Stage:	Final
Protocol version:	3.0, 04-11-2024

Contents

1 Administrative Information 3
1.1 SAP revision history 3
1.2 Administrative details 3
1.3 Roles, responsibilities and approvals 3
2 Aims and Design of the Trial 4
2.1 Background and rationale 4
2.2 Aims and objectives 4
2.3 Trial design 5
2.4 Randomisation, blinding and representativeness of the study 6
2.5 Sample size 6
2.6 Timing of final analysis ..... **Error! Bookmark not defined.**
2.7 Timing of outcome assessments ..... **Error! Bookmark not defined.**
3 Monitoring of the Trial 8
3.1 Operational (logistical) monitoring of patients 8
3.1.1 Screening data 8
3.1.2 Eligibility 8
3.1.3 Recruitment 8
3.1.4 Withdrawal/follow-up 8
3.1.5 Baseline patient characteristics 8
3.1.6 Adherence and protocol deviations 8
3.1.7 Safety data and harms 8
3.1.8 Unblinding 9
3.2 Statistical Monitoring during the trial 9
3.2.1 Assessment of bias 9
3.2.2 Non-compliance / non-adherence 9
3.2.3 Statistical interim analyses and stopping guidance 9
4 Clinical Outcomes and Analysis Data 12
4.1 Primary outcome 12
4.2 Secondary outcomes 13
4.3 Analysis datasets 14

5 Main Statistical Analysis and Estimand Framework 14
5.1 Statistical Principles 14
5.1.1 General considerations 14
5.1.2 Analysis framework: Point estimates and confidence intervals 14
5.2 Estimand Framework 15
5.2.1 Primary estimand 15
5.3 Main Analyses 15
5.3.1 Primary analyses 15
5.3.2 Subgroup analyses 23
5.4 Missing data 23
5.5 Harms 23
5.6 Statistical software 23
6 References 24

1 Administrative Information

1.1 SAP revision history

SAP Version	Protocol Version	Date and timing of revision	Details of revision
1.0	3.0	02-04-2026	N/A

1.2 Administrative details

The study will be analysed with regard to Good Clinical Practice and as laid out in WCTU SOPs. The electronic Study Master File can be found at: M:\WMS\CTU\Cancer\COBIx. This includes the most recent version of the Data Management plan, which gives details of data collection methods.

The Statistical Master File can be found at: <<M drive location>>. All sensitive and confidential material is encrypted and access is limited to the Study Statistician(s) and their nominated access back up (see: M:\WMS\CTU\Statisticians\PGP user tracker for user details).

1.3 Roles, responsibilities and approvals

Role	Name and contact details	Date	Signature
Author of SAP	Dr Peter Kimani Warwick Clinical Trials Unit Peter.K-U.Kimani@warwick.ac.uk	21.04.2026	
Senior statistician	Professor Janet Dunn Warwick Clinical Trials Unit j.a.dunn@warwick.ac.uk		
Chief Investigator	Dr David Snead University Hospitals Coventry and Warwickshire NHS Trust david.snead@uhcw.nhs.uk		
Independent statistician	Professor Nigel Stallard Warwick Clinical Trials Unit N.Stallard@warwick.ac.uk	21 April 2026	

2 Aims and Design of the Study

2.1 Background and 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. However, with the number of pathologists decreasing due to consultant retirement rates outstripping graduates entering the pathology workforce, the increased workload is contributing to delays to diagnosis of cancer and other disease which worsens outcomes for patients (Bainbridge *et al.*, 1016 and Hanna *et al.*, 2020).

This study focusses on endoscopic biopsies from the large bowel that account for 17.8% of the requests in NHS cellular pathology laboratories (unpublished data from seven NHS trusts, Table 1 in the protocol). More than a third (34.5%) of endoscopic biopsies are reported as normal (unpublished data from seven NHS trusts, Table 1 in the protocol). The reporting and diagnosis of endoscopic large bowel biopsies in the United Kingdom is currently performed manually, mostly by consultant pathologists. The aim of this study is to test accuracy of an artificial intelligence (AI) pre-screening tool to screen out normal biopsies to reduce workload and improve workflow. The colon and rectal biopsy (COBIx) algorithm, an AI tool, 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. This would remove normal biopsies from the pathologists' workload and triage abnormal ones, allowing pathologists to focus on abnormal biopsies, 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.

2.2 Aims and objectives

The primary research question is whether the COBIx algorithm can be used to screen out normal biopsies accurately so that pathologists can instead focus their resources on the remaining ones. This will be done by comparing biopsy classifications (normal/abnormal) by pathologists with the same categorisation using the COBIx algorithm.

The secondary research questions are whether there is agreement of categorising abnormal biopsies into

1) neoplastic or non-neoplastic
2) urgent or non-urgent.

2.3 Study design

This multi-site validation study will assess the accuracy of the COBIx algorithm in diagnosing biopsy samples, using pathologists' diagnostic categorisation as the reference standard. Each of the 12 study sites, which are NHS Trusts across the UK, will recruit 100 prospective cases and approximately 800 retrospective cases, resulting into approximately a total of 10,000 cases. A case corresponds to an instance that a patient has biopsies taken. Because samples may be obtained from multiple sites/parts of the colon (bowel), a case can have more than one biopsy. The outcome, diagnostic category, will at biopsy level and so statistical analysis will be at biopsy level. It is expected that the average is three biopsies per case so that the total sample size is expected to be approximately 30,000, with the minimum total sample size being 10,000 (one biopsy per case). Analysis at biopsy level is reasonable because the comparison of the diagnostic categories of a pathologist and the COBIx algorithm is more dependent on whether a biopsy is normal or abnormal than if the biopsies are from the same patient.

Each biopsy will have a single pathologist diagnostic categorisation. Some sites will have multiple study pathologists, but the approach is to share the work so that each case, and hence biopsy, will be reviewed by a single pathologist. For retrospective biopsies, a pathologist will report a diagnostic category for a biopsy from reviewing its existing pathology report. The process for prospective cases is being developed, including deciding the sites to recruit prospective cases and activities for which time taken will be measured. COBIx algorithm will report a single diagnostic category for each biopsy. Both the study pathologists and the COBIx algorithm will categorise each biopsy to one of the following five categories

- Normal: Any biopsy considered to be within normal histological limits or with any subtle abnormalities that are pathologically insignificant.
- Neoplastic urgent: Any neoplastic biopsy that contains any of the following
    - High grade dysplasia
    - Invasive adenocarcinoma
    - Any other malignant tumour
    - Lymphomas (low grade and high grade)
    - Neuroendocrine tumours
    - Spindle cell lesions
- Neoplastic – non-urgent: Any neoplastic biopsy that does not contain any of the above neoplastic features.
- Non-neoplastic urgent: Any non-neoplastic biopsy that contains any of the following
    - Active/ acute inflammation
    - Ulceration
    - Necrosis
    - Ischaemic colitis
- Non-neoplastic non-urgent: Any non-neoplastic entity that does not contain any active inflammation/ ulceration/ necrosis/ ischaemic colitis.

For the primary research question, the last four diagnostic categories are grouped into a single group "Abnormal". Combining pathologists and COBIx algorithm diagnostic categories, data to answer the primary research question will be summarised as in Table 1.

For the secondary research questions, two other recategorizations will be used, that is urgent versus non-urgent and neoplasia versus non-neoplasia. The data will also be summarised using two-by-two tables, similar to Table 1.

Table 1: Summary of primary research question data

COBIx algorithm (New diagnostic tool)	Pathologists' diagnostic category (Reference standard)		Total
	Disease (Abnormal)	No disease (Normal)	
Disease (Abnormal)	TP	FP	TP+FP
No disease (Normal)	FN	TN	FN+TN
Total	TP+FN	FP+TN	TP+FP+FN+TN

TP = True positive; FP = False positive; FN = False negative; TN = True negative.

2.4 Randomisation, blinding and representativeness of the study

For each biopsy, both a pathologist diagnostic category and a COBIx algorithm diagnostic category will be reported. There is no randomisation to the order; COBIx algorithm diagnostic categories will be reported after pathologists' diagnoses. However, there is no risk of bias because the team running the COBIx algorithm will only receive whole slide images of the biopsies and their IDs, and they will be blind to pathologists' diagnostic categories.

The study will recruit retrospective and prospective pathology cases with an approximate ratio of 8:1. For the retrospective cases, sites have been directed that 90% are routine cases, representing the usual day to day cases seen in their laboratory, with the remaining 10% being rare and unusual entities. The enrichment is so that study results can be generalised to include rare and unusual entities.

We do not expect reporting bias by study pathologists because diagnoses for retrospective cases are available on existing pathology notes/reports, with study pathologists only needing to extract the diagnoses from the notes and allocating to one of the five diagnostic categories.

The process for the prospective cases is not finalised. Randomisation, blinding and representativeness of these cases will be determined when the process is finalised.

2.5 Sample size

The study aims to enrol at least 10,000 histopathology cases. Sample size for the analysis corresponds to the number of biopsies and, on average, it is expected that each histopathology case will have three biopsies. However, in assessing sample size adequacy, we were conservative assuming that there will be one biopsy per case. A sample size of 10,000 is adequate to estimate diagnostic accuracy measures (sensitivity, specificity, negative predictive value (NPV) and positive predictive value (PPV)) with high precision. There is no hypothesis testing with the sole aim being to

estimate the diagnostic accuracy of the COBIx algorithm with high precision. However, to determine the interim analysis threshold in Section 3.2.3, among other considerations such as what the false negatives are (e.g. are they missed cancers?), it was discussed that the algorithm can only be adopted in practice if sensitivity is at least 0.95 and NPV is at least 0.90. Therefore, we assessed precision for NPV=0.90. Assuming we will observe NPV=0.90 and the same number of normal cases by pathologists and COBIx algorithm, varying prevalence between 49% and 90% (data from seven sites in Table 1 in the protocol), the largest margin of error (1.96 x standard error) for any of the four diagnostic measures is 0.0186, ensuring very narrow confidence intervals for all the diagnostic measures. The calculation when prevalence is 90% is presented in Box 1. It is essential to deliver comprehensive assessment of the COBIx tool across all patient groups and the full range of pathological entities encountered in practice and this sample size is deemed adequate to achieve this.

Box 1: Assessing the margin of error for one scenario considered in assessing sample size adequacy

Fixed quantities

- Total sample size = 10000.
- Prevalence = 90% so that 1000 normal for pathologists (reference standard).
- COBIx algorithm normal = pathologist normal, so 1000.
- Observed NPV = 0.9.
- With these restrictions, the observed data is

	Pathologist Category		
COBIx	**Abnormal**	**Normal**	**Total**
Abnormal	8900	100	9000
Normal	100	900	1000
Total	9000	1000	10000

- Sensitivity = 0.989 (8900/9000); SE = $\sqrt{Sensitivity \times (1 - Sensitivity)/9000} = 0.0011$; ME = 1.96SE = 0.0022 so that 95% CI =(0.989-ME, 0.989+ME) = (0.987,0.991). Note small ME so that the confidence interval is narrow.
- Specificity = 0.900 (900/1000); SE = $\sqrt{Sensitivity \times (1 - Sensitivity)/1000} = 0.0095$; ME = 1.96SE = 0.0186 so that 95% CI = (0.900-ME, 0.900+ME) = (0.881, 0.919). Note small ME so that the confidence interval is narrow.
- Calculations for PPV give same results as sensitivity and calculations for NPV give same results.
- Other scenarios changed the prevalence from 90% to other values between 49% and 90% and the largest ME for a diagnostic accuracy measure was 0.0186. The one observed in the scenario presented in this box.

3 Monitoring of the Study

No patients are recruited into this study. Most histopathology cases recruited are retrospective, with pathologists' diagnostic categories extracted from patient records. COBIx diagnostic categories will also be read rapidly. Therefore, comparison data are obtained rapidly and thus there is little scope for monitoring during the study. At interim analysis, FNs and FPs will be reviewed by study pathologists to determine if there are patient safety issues to be addressed.

3.1 Operational (logistical) monitoring of patients

3.1.1 Screening data

Histopathology cases will be retrospectively (~90% of them) recruited consecutively from laboratory records. Therefore, there is no risk to representativeness.

3.1.2 Eligibility

Study pathologists will recruit histopathology cases of endoscopic biopsies from the large bowel. There is negligible risk of including biopsies that are not from the large bowel and this would be picked by COBIx algorithm .

3.1.3 Recruitment

A Consolidated Standards of Reporting Trials (CONSORT) flow diagram is not relevant to this study. All information for eligible cases will be recorded. It is not be logical to collect data for ineligible cases.

3.1.4 Withdrawal/follow-up

This is not applicable for this study because patients are not followed up.

3.1.5 Baseline patient characteristics

A characteristics table will present descriptive statistics summarising the patients and tumours included as cases in the study.

3.1.6 Adherence and protocol deviations

This does not apply because the COBIx study is not an interventional study.

3.1.7 Safety data and harms

At interim analysis, FNs and FPs will be reviewed by study pathologists to determine if there are patient safety issues to be addressed. This will be repeated at final analysis.

3.1.8 Unblinding

This does not apply because patients are not recruited and the type of the study.

3.2 Statistical Monitoring during the study

3.2.1 Assessment of bias

There is no comparison of groups so comparison of characteristics between groups does not apply. The only bias that could arise is selection bias with the study sample not being a representative of the target population. Histopathology cases are recruited consecutively to try and avoid the selection bias.

3.2.2 Non-compliance / non-adherence

This does not apply because the COBIx study is not an interventional study.

3.2.3 Statistical interim analyses and stopping guidance

An interim analysis to assess if the COBIx algorithm should be redeveloped (modify) or not (do not modify) will be performed using approximately the first 40% histopathology cases recruited from each site. The percentage is approximate because 40% will be based on the number of expected cases to be recruited at each site at the point of determining the interim analysis cases. We emphasize that the percentage of interim analysis cases will be independent of the accruing study data comparing pathologists' diagnoses to COBIx algorithm's diagnoses. The analysis described in Section 5.3.1 remains valid as long as the percentage of cases used in the interim analysis is independent of data comparing pathologists' diagnoses to COBIx algorithm's diagnoses, which is the case in this study. The decision rule used to decide whether to modify the COBIx algorithm or not is described in detail next.

Assuming the study recruits the exact target number of histopathology cases per site size and interim analysis uses exactly 40% of the cases, the study design is summarised in Figure 1. Stage 1 consists of recruiting the first 4,000/10,000 cases (approximately 12,000 biopsies), obtaining both the pathologists' and COBIx algorithm diagnostic categories data for all the biopsies and performing an interim analysis. The interim analysis will entail computing sensitivity, specificity, NPV and PPV using the stage 1 data. Stage 2 consists of recruiting the remaining 6,000/10,000 cases (approximately 18,000 biopsies), obtaining the pathologists' and COBIx algorithm diagnostic

categories data for the biopsies and performing the end of study final analysis. The realised study numbers may be different.

The version of the COBIx algorithm to be used to determine diagnostic categories for biopsies in stage 2 and the end of study final analysis will depend on the interim analysis results. Based on stage 1 findings, **the COBIx algorithm will be modified if** sensitivity is below 0.94 or NPV is below 0.88. A simulation study that informed this decision boundary is described in Appendix 1. Modification will involve using stage 1 data to redevelop the COBIx algorithm. The modified/redeveloped algorithm will be validated with stage 2 data only. That is, at the end of the study, the primary analysis to compute diagnostic measures will use stage 2 data only.

Using stage 1 data, the COBIx algorithm will not be modified if sensitivity is at least 0.94 and NPV is at least 0.88. Because, in this case, the same algorithm is used to determine diagnostic categories of biopsies in stage 1 and in stage 2, at the end of the study, the COBIx algorithm will be validated using both stage 1 and stage 2 data. However, there would need to be statistical adjustment at the final analysis because the decision to not modify the algorithm would have been based on having high sensitivity value (at least 0.94) and a high NPV value (at least 0.88) in stage 1, which may lead to positive bias. Therefore, for the analysis corresponding to the primary research question, building on Robertson et al. (2016) and Khan et al. (2023) we have developed new estimators that will be used to compute unbiased estimates for sensitivity, specificity, NPV and PPV. The estimators are given in Section 5.3.1.5, and the independent study steering committee (SSC) statistician has seen the manuscript of the work developing the estimators. The SAP has also been reviewed by the CTU director as the independent statistician familiar with adaptive designs. To compute 95% confidence intervals, building on Pepe et al. (2009) and based on estimates computed using estimators developed for this study, a bootstrap algorithm that adjusts for the possibility of redeveloping the algorithm after interim analysis will be used. The bootstrap algorithm is given in Section 5.3.1.6.

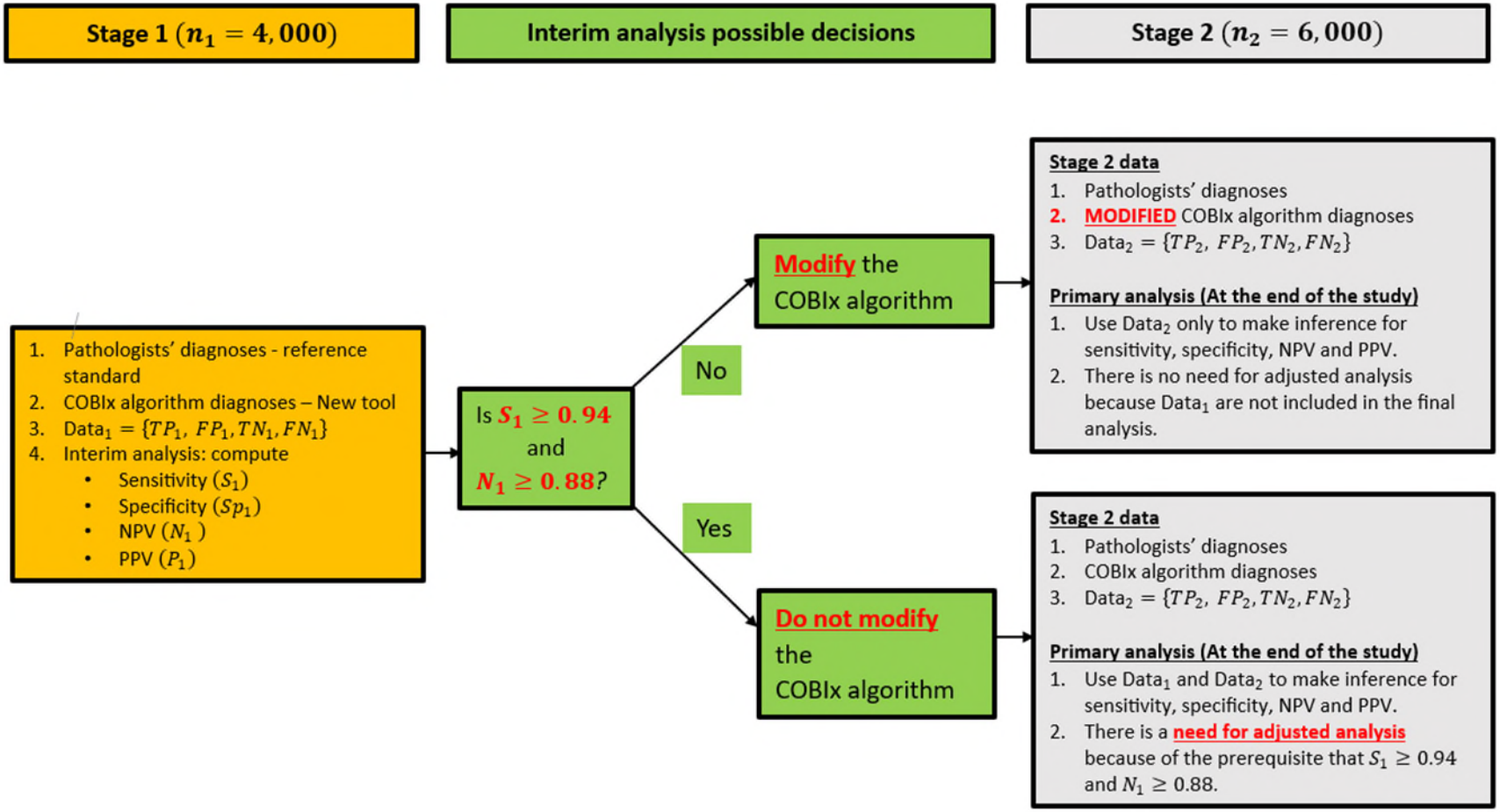

Figure 1: A summary of the COBIx study design. Sample size corresponds to number of biopsies. Subscripts are added to diagnostic measures to distinguish measures computed with stage 1 data only (subscript is 1) and those computed with stage data only (subscript is 2).

4 Clinical Outcomes and Analysis Data

4.1 Primary outcome

For the primary research question, a pathologist's diagnostic category for a biopsy will be compared to the COBIx algorithm's diagnostic category. The outcome for each comparison will be either true positive (TP) if both a pathologist's and COBIx algorithm's diagnostic categories indicate an abnormal biopsy, false positive (FP) if a pathologist's diagnostic category indicates a normal biopsy whilst COBIx algorithm's diagnostic category indicates an abnormal biopsy, false negative (FN) if a pathologist's diagnostic category is an abnormal biopsy whilst COBIx algorithm's diagnostic category indicates a normal biopsy, or true negative (TN) if both a pathologist's and COBIx algorithm diagnostic categories indicate a normal biopsy. These outcome data will be used to compute diagnostic measures sensitivity, specificity, negative predictive value (NPV) and positive predictive value (PPV). Using notation in Table 1, where cells correspond to frequencies (count) of the different outcomes, the diagnostic measures are defined as:

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{Negative predictive value (NPV)} = \frac{TN}{TN + FN}$$

$$\text{Positive predictive value (PPV)} = \frac{TP}{TP + FP}.$$

Reporting the diagnostic measures will be accompanied by the interpretation in Table 2(a). The prevalence of abnormal cases (based on reference standard) will also be reported because NPV and PPV can be computed from sensitivity, specificity and prevalence. Also, incorporating the study's observed diagnostic accuracy measures and observed prevalence, categorisation for 1,000 biopsies will be presented as in Table 2(b).

Table 2(a): Interpretation of results (illustration with made up data)

Diagnostic measure	Possible results†	Interpretation
Sensitivity	0.951	For every 1,000 abnormal biopsies, COBIx algorithm will incorrectly categorise 49 of them [$= 1000 \times (1 - 0.951)$] as normal. That is, 49 false negatives per 1,000 abnormal biopsies.
Specificity	0.900	For every 1,000 normal biopsies, COBIx algorithm will incorrectly categorise 100 of them [$=1000 \times (1 - 0.900)$] as abnormal. That is, 100 false positives per 1,000 normal biopsies.
NPV	0.908	For every 1,000 biopsies categorised as normal by the COBIx algorithm, 92 [$=1000 \times (1 - 0.908)$] are in fact abnormal. That is, 92 are false negatives.
PPV	0.946	For every 1,000 biopsies categorised as abnormal by the COBIx algorithm, 54 [$=1000 \times (1 - 0.946)$] are in fact normal. That is, 54 are false positives.

†These are made up results to illustrate interpretation that we will give when reporting diagnostic measure values.

Table 2(b): Illustration of categorisation of 1000 biopsies based on the made up diagnostic measures in Table 2(a) and expected study prevalence of abnormal cases of 0.65

	Pathologist category	
COBIx algorithm category	**Abnormal**	**Normal**
Abnormal	618	35
Normal	32	315

4.2 Secondary outcomes

For the secondary research questions, other categorisations neoplastic versus non-neoplastic (secondary research question 1) and urgent versus non-urgent (secondary research question 2), will be used to determine accuracy of the COBIx algorithm. In these categorisations, a biopsy categorised normal in Table 1 is non-neoplastic for secondary research question 1 and non-urgent for secondary research question 2. The data will be summarised as shown in Table 3(a) (using example of urgent versus non-urgent categorisation) and the diagnostic accuracy measures used in the primary analysis (sensitivity, specificity, NPV and PPV) will be used to analyse them. The analysis will be repeated using only a subset of the biopsies that are categorised abnormal by the pathologists.

Table 3(a): Summary of secondary research question data

	Pathologist diagnosis	
COBIx algorithm	Urgent	Non-urgent
Urgent	"TP"	"FP"
Non-urgent	"FN"	"TN"

Secondary analysis will involve analysing the data based on three categories (Table 3(b)), reporting for each column and each row, the proportion in each cell.

Table 3(b): Summary of secondary research question data

COBIx algorithm	Pathologist diagnosis		
	Urgent	Non-urgent	Normal
Urgent	TU	FU	FP(1)
Non-urgent	FN-U	TN-U	FP(2)
Normal	FN(1)	FN(2)	TN

TU = True urgent; TN-U = True non-urgent; FU = False urgent; FN-U = False non-urgent

4.3 Analysis datasets

A single dataset will be analysed. Missing data is not anticipated. All pathologists' diagnostic categories are available and the only reason that COBIx diagnostic category could be unavailable is if the scanned image is damaged. If there are biopsies that COBIx is unable to report diagnostic categories for, these will be summarised descriptively and not included in the main analysis.

5 Main Statistical Analysis and Estimand Framework

5.1 Statistical Principles

5.1.1 General considerations

Outcome data will be summarised as in Table 1, Table 3(a) and Table 3(b). Characteristics of the biopsies will be summarised as described in Section 5.3.1.1.

5.1.2 Analysis framework: Point estimates and confidence intervals

The primary analysis will entail estimating four diagnostic measures (sensitivity, specificity, NPV and PPV) and their corresponding 95% confidence intervals (CIs). There will be no hypothesis testing. A positive study outcome, corresponding to the primary research question, will be that the COBIx algorithm is a successful diagnostic tool for screening out biopsies without disease (normal biopsies). Because the desire is to screen out normal biopsies whilst, for patient safety, minimising false negatives, this corresponds to high sensitivity and NPV. During protocol development thresholds were set to be 0.95 for sensitivity and 0.90 for NPV. The thresholds were informed by the results of INGUANA algorithm (Graham et al., 2023), considering cases mix (e.g. cancer) and prevalence of disease. However, sensitivity and NPV will not be the exclusive metrics to determine the success of the COBIx algorithm. Specificity, PPV, workload reduction and in depth analysis of false negatives will also be interpreted. The COBIx algorithm will be built before the interim analysis by refining the INGUANA algorithm using independent data that will not be part of the interim analysis data.

5.2 Estimand Framework

5.2.1 Primary estimand

Details of the primary estimand are summarised in the table below.

Primary estimand and its associated estimator

Estimand attributes	Estimator
Objective To assess diagnostic accuracy of the COBIx algorithm in providing diagnostic categories for endoscopic biopsies from the large bowel when compared to pathologists' diagnostic categories.	
Treatment conditions There is no comparison of treatments. For each endoscopic biopsy, the diagnostic category of a pathologist is compared to the diagnostic category of the COBIx algorithm.	
Target population Endoscopic biopsies from the large bowel received in NHS cellular pathology laboratories in the UK.	*Analysis set* Analysis set consists of histopathology cases (a case can have multiple biopsies), with most cases retrospectively recruited sequentially as received in the NHS cellular pathology laboratories within a time period chosen by the enrolling site.
Variable (primary outcome) *The outcome is the result of the comparison of the diagnostic categories of the COBIx algorithm and of the pathologists. With pathologist's category being the reference standard, the possible outcomes of a comparison are* true positive (TP), false positive (FP), false negative (FN) or true negative (TN).	*Measurement/endpoint obtained to address objective* Whether when COBIx algorithm diagnostic category is compared to a pathologist's diagnostic category, it is a TP, FP, FN or TN.
Summary measure Diagnostic accuracy measures sensitivity, specificity, negative predictive value (NPV) and positive predictive value (PPV)	*Analysis approach* As described in Section 5.1.2, the analysis entails computing point estimates and 95% confidence intervals for diagnostic accuracy measures.

5.3 Main Analyses

5.3.1 Primary analyses

5.3.1.1 Descriptive analysis

Characteristics of the patients, cases and biopsies will be summarised using descriptive statistics. For categorical characteristics, count and percentage in categories will be presented and for continuous characteristics, the minimum, maximum, median (Lower quartile, upper quartile) and mean (standard deviation) will be reported. Exploratory comparison of pathologists diagnostic categories and COBIx algorithm diagnostic categories will be summarised using tables using the format of Table 1, Table 2, Table 3(a) and Table 3(b).

Descriptive analysis will be performed using all data and by the subgroups described above.

5.3.1.2 Reporting diagnostic measures

As described in Section 5.1.2, the primary analysis will estimate diagnostic accuracy measures (including 95% confidence intervals (CIs)) to assess accuracy of the COBIx algorithm in diagnosing colon and rectal endoscopic biopsies. The usual practice is to report diagnostic measures in two pairs (1) sensitivity and specificity, and (2) NPV and PPV. The same format will be used in main texts of reports and publications whilst in tables, the order will be sensitivity, specificity, NPV and PPV. All results (estimates and 95% confidence intervals (CIs)) will be reported at three decimal places.

To aid translation of the results to a typical clinical population, we will have subgroup analysis, computing diagnostic accuracy measures and 95% CIs for each category of the variables listed in Section 5.3.2. Also, diagnostic accuracy measures and 95% CIs for a typical clinical population will be computed as the weighted mean of diagnostic measures of different categories, using appropriate clinical population category prevalences as the weights, not our sample prevalences (e.g. there is study sample enriching to include minority ethnicities at a proportion higher than the clinical population). Adjusted diagnostic measures for subgroups is possible with 95% CIs computed using a bootstrap algorithm like the one for overall analysis that is described in Section 5.3.1.6. Typical clinical population diagnostic accuracy measures calculation will depend on whether reasonable estimates of prevalences of factors of a variable(s) in Section 5.3.2 can be determined.

After describing in Section 5.3.1.3 how the study data will be summarised and the notation used for the summaries, depending on interim analysis decision, it is described in subsequent sections (5.3.1.4 to 5.3.1.6) how the four diagnostic measures and the corresponding 95% CIs will be computed.

5.3.1.3 Stagewise data

For the primary research question, Table 4 shows how data collected in each stage of the study and data for the combined stage 1 and stage 2 will be summarised.

Table 4: Summary of primary data in different stages

COBIx algorithm	Stage 1 data			Stage 2 data			Combined stage 1 and stage 2 data†		
	Reference standard			Reference standard			Reference standard		
	Disease	No disease	Total	Disease	No disease	Total	Disease	No disease	Total
Disease	A_1 (TP)	B_1 (FP)	$A_1 + B_1$	A_2 (TP)	B_2 (FP)	$A_2 + B_2$	$A = A_1 + A_2$ (TP)	$B = B_1 + B_2$ (FP)	$A + B$
No disease	C_1 (FN)	D_1 (TN)	$C_1 + D_1$	C_2 (FN)	D_2 (TN)	$C_2 + D_2$	$C = C_1 + C_2$ (FN)	$D = D_1 + D_2$ (TN)	$C + D$
Total	$A_1 + C_1$	$B_1 + D_1$	n_1	$A_2 + C_2$	$B_2 + D_2$	n_2	$A + C$	$B + D$	$n = n_1 + n_2$

Stage 1 and stage 2 sample sizes are $n_1 = A_1 + B_1 + C_1 + D_1$ and $n_2 = A_2 + B_2 + C_2 + D_2$, respectively, and the total sample size is $n = n_1 + n_2$.

†Aggregation is only valid if the decision at interim analysis is "do not modify" so that numbers are based on the same diagnostic tool.

Table 5: Summary of formulae for computing diagnostic measures with different sets of data

Diagnostic measure	Stage 1		Stage 2		Combined stage 1 and stage 2†	
	Estimate	Standard error (SE)	Estimate	Standard error (SE)	Estimate	Standard error (SE)
Sensitivity	$\hat{S}_1 = \frac{A_1}{A_1 + C_1}$	$\sigma_{1,S} = \sqrt{\frac{\hat{S}_1(1-\hat{S}_1)}{A_1 + C_1}}$	$\hat{S}_2 = \frac{A_2}{A_2 + C_2}$	$\sigma_{2,S} = \sqrt{\frac{\hat{S}_2(1-\hat{S}_2)}{A_2 + C_2}}$	$\hat{S} = \frac{A}{A + C}$	$\sigma_S = \sqrt{\frac{\hat{S}(1-\hat{S})}{A + C}}$
Specificity	$\hat{S}p_1 = \frac{D_1}{B_1 + D_1}$	$\sigma_{1,Sp} = \sqrt{\frac{\hat{S}_{p1}(1-\hat{S}_{p1})}{B_1 + D_1}}$	$\hat{S}p_2 = \frac{D_2}{B_2 + D_2}$	$\sigma_{2,Sp} = \sqrt{\frac{\hat{S}_{p2}(1-\hat{S}_{p2})}{B_2 + D_2}}$	$\hat{S}p = \frac{D}{B + D}$	$\sigma_{Sp} = \sqrt{\frac{\hat{S}p(1-\hat{S}p)}{B + D}}$
NPV	$\widehat{N}_1 = \frac{D_1}{C_1 + D_1}$	$\sigma_{1,N} = \sqrt{\frac{\widehat{N}_1(1-\widehat{N}_1)}{C_1 + D_1}}$	$\widehat{N}_2 = \frac{D_2}{C_2 + D_2}$	$\sigma_{2,N} = \sqrt{\frac{\widehat{N}_2(1-\widehat{N}_2)}{C_2 + D_2}}$	$\widehat{N} = \frac{D}{C + D}$	$\sigma_N = \sqrt{\frac{\widehat{N}(1-\widehat{N})}{C + D}}$
PPV	$\hat{P}_1 = \frac{A_1}{A_1 + B_1}$	$\sigma_{1,P} = \sqrt{\frac{\hat{P}_1(1-\hat{P}_1)}{A_1 + B_1}}$	$\hat{P}_2 = \frac{A_2}{A_2 + B_2}$	$\sigma_{2,P} = \sqrt{\frac{\hat{P}_2(1-\hat{P}_2)}{A_2 + B_2}}$	$\hat{P} = \frac{A}{A + B}$	$\sigma_P = \sqrt{\frac{\hat{P}(1-\hat{P})}{A + B}}$

NPV = Negative predictive value; PPV = Positive predictive value.

†Aggregation is only valid if the decision at interim analysis is "do not modify" so that numbers are based on the same diagnostic tool.

Using the definitions in Section 4.1, estimates for diagnostic accuracy measures using stage 1 data only, stage 2 data only and using the combined stage 1 and stage 2 data can be obtained as shown in Table 5. The expressions for standard errors (SEs) for the estimates are also given in Table 5. The SEs are based on the normal approximation to the binomial distribution. We anticipate the approximation to work well because of the large sample size.

5.3.1.4 The case of modifying the COBIx algorithm after interim analysis: Analysis for the primary research question

If the COBIx algorithm is modified after the interim analysis (based on stage 1 data), stage 2 estimates given in Table 5 will be used to independently validate accuracy of the modified COBIx algorithm. The expressions for the estimates are those in the column under the header "Stage 2" and "Estimate". For example, the estimate for sensitivity will be computed as follows

$$\text{Sensitivity}\left(\hat{S}_2\right) = \frac{A_2}{A_2 + C_2}, \qquad (1)$$

where A_2 and C_2 respectively denote true positives and false negatives from the data collected in stage 2.

The 95% confidence interval for each diagnostic measure will be obtained using the expression,

$$\text{Estimate} \pm [1.96 \times \text{SE}], \qquad (2)$$

where the expressions for the standard errors (SE) are those in Table 5, the column under the header "Stage 2" and "Standard error (SE)". For example, for sensitivity, the SE is given by

$$\sigma_{2,S} = \sqrt{\frac{\hat{S}_2(1-\hat{S}_2)}{A_2 + C_2}}.$$

These estimates, e.g. expression (1) for sensitivity, and the 95% confidence intervals obtained using expression (2) will be the primary analysis results if the COBIx algorithm is modified after the interim analysis. The results will be summarised as in Table 6.

Table 6: Template table for summarising results if the COBIx algorithm is modified after the interim analysis

Diagnostic measure	**Estimate (95% confidence interval)**
Sensitivity Specificity Negative predictive value (NPV) Positive predictive value (PPV)	

5.3.1.5 The case of not modifying the COBIx algorithm after the interim analysis: Analysis for the primary research question – point estimates

If the COBIx algorithm is not modified, the estimates in Table 5 that are obtained using the combined stage 1 and stage 2 data (second last column) can overestimate the accuracy of the COBIx algorithm. This is because stage 1 data have been proven to produce sensitivity and NPV at or above the pre-specified decision thresholds of 0.94 and 0.88, respectively. We refer to such estimates that do not adjust for not modifying the COBIx algorithm because of observing high sensitivity and NPV values at interim analysis as the naïve estimates.

In the final analysis performed at the end of the study, unbiased estimates for the diagnostic measures (sensitivity, specificity, NPV and PPV) that adjust for not modifying the COBIx algorithm because, at interim analysis, sensitivity ≥ 0.94 and NPV ≥ 0.88, will be computed using adjusted estimators that have been derived to analyse this study. New estimators have been derived because existing estimators that adjust for interim analysis decision do not account for the correlation between stage 1 estimators. To demonstrate correlation, for example, if sensitivity estimate is 1, NPV estimate is also 1 and hence sensitivity and NPV estimates are correlated. Building on Robertson *et al.* (2016) and Khan *et al.* (2023), we have developed new estimators that through simulations have been shown to have better properties than the naïve estimators and existing estimators that adjust for decision rule but ignore the fact that stage 1 estimates are correlated. The work on developing the new estimators has been written as a manuscript to be submitted to a journal soon. The manuscript has been shared with the independent SSC statistician.

To obtain the expressions that will be used to compute unbiased estimates for sensitivity, specificity, NPV and PPV at the end of the study, the joint density of stage 1 sensitivity, specificity, NPV and PPV is required. The asymptotic multivariate normal distribution is assumed, and the variances are the squares of the stage 1 standard errors given in Table 5. It is known that the correlation between sensitivity and specificity is zero. Bootstrap samples will be used to get the values for the other pairwise correlations between estimators for sensitivity, specificity, NPV and PPV. The analysis is at the end of the study so that bootstrap samples can be drawn using information from stage 1 and stage 2 data. Notation for the correlations of stage 1 diagnostic estimators is summarised in Table 7. As in Table 5, let $\hat{S}_2$ and $\hat{S}p_2$ denote stage 2 estimates for sensitivity and specificity, respectively, while, based on diagnoses by the pathologists, let n_1^+ and n_1^- denote stage 1 number of abnormal biopsies and number of normal biopsies, respectively. Bootstrap stage 1 TP and TN are simulated assuming $\text{Binomial}(\hat{S}_2, n_1^+)$ and $\text{Binomial}(\hat{S}p_2, n_1^-)$, respectively. Stage 2 estimates $\hat{S}_2$ and $\hat{S}p_2$ are used because they are unbiased and hence expected to give more accurate estimates for correlations compared to using the stage 1 estimators $\hat{S}_1$ and $\hat{S}p_1$ that could be biased. 5000 bootstrap samples will be simulated and for each simulated sample, sensitivity, specificity, NPV and PPV estimates will be computed. Pairwise correlations for the estimates from the 5000 bootstrap samples are the correlations for stage 1 diagnostic estimators. For a fixed sample size, the correlations are positive and so negative correlations will be fixed to be zero.

Table 7: Notation for correlation of stage 1 diagnostic estimators

Correlation notation	$\rho_{S,Sp}$	$\rho_{S,N}=0$	$\rho_{S,P}$	$\rho_{Sp,N}$	$\rho_{Sp,P}$	$\rho_{N,P}$
Pair of stage 1 estimators	$\hat{S}_1$ and $\hat{S}p_1$	$\hat{S}_1$ and $\hat{N}_1$	$\hat{S}_1$ and $\hat{P}_1$	$\hat{S}p_1$ and $\hat{N}_1$	$\hat{S}p_1$ and $\hat{P}_1$	$\hat{N}_1$ and $\hat{P}_1$

$\hat{S}_1$, $\hat{S}p_1$, $\hat{N}_1$ and $\hat{P}_1$ denote stage 1 estimators for sensitivity, specificity, NPV and PPV, respectively.

The expressions to compute the unbiased estimates for the sensitivity, specificity, NPV and PPV are given below.

- Sensitivity

Let $\hat{S}_1$, $\hat{S}_2$, $\hat{N}_1$, $\sigma_{1,S}$, $\sigma_{2,S}$, $\sigma_{1,N}$ and $\hat{S}$ be as defined in Table 5, $\rho_{S,N}$ be as defined as in Table 7 and F_1 and F_2 denote the futility boundaries for sensitivity and NPV, respectively. Define $\hat{\theta}^*_{S,S}=\hat{S}_1+\frac{\sigma^2_{1,S}}{\sigma^2_{2,S}}\hat{S}_2$, $U^1_S=\frac{\sigma^2_{2,S}}{\sigma^2_{1,S}}\left(\hat{\theta}^*_{S,S}-F_1\right)$, $\hat{\theta}^*_{N,S}=\hat{N}_1+\frac{\rho_{S,N}\sigma_{1,S}\sigma_{1,N}}{\sigma^2_{2,S}}\hat{S}_2$, $U^2_S=\frac{\sigma^2_{2,S}}{\rho_{S,N}\sigma_{1,S}\sigma_{1,N}}\left(\hat{\theta}^*_{N,S}-F_2\right)$ and $U_S=\min\{U^1_S,U^2_S\}$. Further let ϕ and Φ denote the density function of standard normal and distribution function of standard normal, respectively, and define $W^{(U)}_S=\frac{\sqrt{\sigma^2_{1,S}+\sigma^2_{2,S}}}{\sigma^2_{1,S}}\left(U_S-\frac{\sigma^2_{2,S}\hat{S}_1+\sigma^2_{1,S}\hat{S}_2}{\sigma^2_{1,S}+\sigma^2_{2,S}}\right)$. The asymptotically uniformly minimum variance conditional unbiased estimator (UMVCUE) for sensitivity is given by

$$\hat{\theta}_{S,UMVCUE}=\hat{S}-\frac{\sigma^2_{2,S}}{\sqrt{\sigma^2_{1,S}+\sigma^2_{2,S}}}\frac{\phi\left(W^{(U)}_S\right)}{\Phi\left(W^{(U)}_S\right)}. \quad (3)$$

- Specificity

Let $\hat{S}p_1$, $\hat{S}p_2$, $\hat{S}_1$, $\hat{N}_1$, $\sigma_{1,Sp}$, $\sigma_{2,Sp}$, $\sigma_{1,S}$, $\sigma_{1,N}$ and $\hat{S}p$ be as defined in Table 5, $\rho_{Sp,N}$ be as defined as in Table 7 and F_2 denote the futility boundaries for NPV. Define $\hat{\theta}^*_{N,Sp}=\hat{N}_1+\frac{\rho_{Sp,N}\sigma_{1,Sp}\sigma_{1,N}}{\sigma^2_{2,Sp}}\hat{S}p_2$, $U_{Sp}=\frac{\sigma^2_{2,Sp}}{\rho_{Sp,N}\sigma_{1,S_p}\sigma_{1,N}}\left(\hat{\theta}^*_{N,Sp}-F_2\right)$. Further define $W^{(U)}_{Sp}=\frac{\sqrt{\sigma^2_{1,Sp}+\sigma^2_{2,Sp}}}{\sigma^2_{1,Sp}}\left(U_{Sp}-\frac{\sigma^2_{2,Sp}\hat{S}_{p1}+\sigma^2_{1,Sp}\hat{S}_{p2}}{\sigma^2_{1,S_p}+\sigma^2_{2,S_p}}\right)$. The asymptotically UMVCUE for specificity is given by

$$\hat{\theta}_{Sp,UMVCUE}=\hat{S}p-\frac{\sigma^2_{2,Sp}}{\sqrt{\sigma^2_{1,Sp}+\sigma^2_{2,Sp}}}\frac{\phi\left(W^{(U)}_{Sp}\right)}{\Phi\left(W^{(U)}_{Sp}\right)}. \quad (4)$$

- Negative predictive value (NPV)

Let $\widehat{N}_1$, $\widehat{N}_2$, $\hat{S}_1$, $\sigma_{1,N}$, $\sigma_{2,N}$, $\sigma_{1,S}$ and $\widehat{N}$ be as defined in Table 5, $\rho_{S,N}$ be as defined as in Table 7 and F_1 and F_2 denote the futility boundaries for sensitivity and NPV, respectively. Define $\hat{\theta}^*_{S,N} = \hat{S}_1 + \frac{\rho_{S,N}\sigma_{1,S}\sigma_{1,N}}{\sigma^2_{2,N}}\widehat{N}_2$, $U^1_N = \frac{\sigma^2_{2,N}}{\rho_{S,N}\sigma_{1,S}\sigma_{1,N}}\left(\hat{\theta}^*_{S,N} - F_1\right)$, $\hat{\theta}^*_{N,N} = \widehat{N}_1 + \frac{\sigma^2_{1,N}}{\sigma^2_{2,N}}\widehat{N}_2$, $U^2_N = \frac{\sigma^2_{2,N}}{\sigma^2_{1,N}}\left(\hat{\theta}^*_{N,N} - F_2\right)$ and $U_N = \min\{U^1_N, U^2_N\}$. Further define $W^{(U)}_N = \frac{\sqrt{\sigma^2_{1,N}+\sigma^2_{2,N}}}{\sigma^2_{1,N}}\left(U_U - \frac{\sigma^2_{2,N}\widehat{N}_1+\sigma^2_{1,N}\widehat{N}_2}{\sigma^2_{1,N}+\sigma^2_{2,N}}\right)$. The asymptotically UMVCUE for NPV is given by

$$\hat{\theta}_{N,UMVCUE} = \widehat{N} - \frac{\sigma^2_{2,N}}{\sqrt{\sigma^2_{1,N}+\sigma^2_{2,N}}}\frac{\phi\left(W^{(U)}_N\right)}{\Phi\left(W^{(U)}_N\right)}. \quad (5)$$

- Positive predictive value (PPV)

Let $\hat{P}_1$, $\hat{P}_2$, $\hat{S}_1$, $\widehat{N}_1$, $\sigma_{1,P}$, $\sigma_{2,P}$, $\sigma_{1,S}$, $\sigma_{1,N}$ and $\hat{P}$ be as defined in Table 5, $\rho_{S,P}$ and $\rho_{N,P}$ be as defined as in Table 7 and F_1 and F_2 denote the futility boundaries for sensitivity and NPV, respectively. Define $\hat{\theta}^*_{S,P} = \hat{S}_1 + \frac{\rho_{S,P}\sigma_{1,S}\sigma_{1,P}}{\sigma^2_{2,P}}\hat{P}_2$, $U^1_P = \frac{\sigma^2_{2,P}}{\rho_{S,P}\sigma_{1,S}\sigma_{1,P}}\left(\hat{\theta}^*_{S,P} - F_1\right)$,, $\hat{\theta}^*_{N,P} = \widehat{N}_1 + \frac{\rho_{N,P}\sigma_{1,P}\sigma_{1,N}}{\sigma^2_{2,P}}\hat{P}_1$, $U^2_P = \frac{\sigma^2_{2,P}}{\rho_{N,P}\sigma_{1,P}\sigma_{1,N}}\left(\hat{\theta}^*_{N,P} - F_2\right)$ and $U_P = \min\{U^1_P, U^2_P\}$. Further define $W^{(U)}_P = \frac{\sqrt{\sigma^2_{1,P}+\sigma^2_{2,P}}}{\sigma^2_{1,P}}\left(U_P - \frac{\sigma^2_{2,P}\hat{P}_1+\sigma^2_{1,P}\hat{P}_2}{\sigma^2_{1,P}+\sigma^2_{2,P}}\right)$. The asymptotically UMVCUE for PPV is given by

$$\hat{\theta}_{P,UMVCUE} = \hat{P} - \frac{\sigma^2_{2,P}}{\sqrt{\sigma^2_{1,P}+\sigma^2_{2,P}}}\frac{\phi\left(W^{(U)}_P\right)}{\Phi\left(W^{(U)}_P\right)}. \quad (6)$$

The primary analysis will be the conditionally unbiased estimates. Naïve estimates will be reported for sensitivity analysis. The R code that will be used to compute various estimates is provided in Appendix 2.

5.3.1.6 The case of not modifying the COBIx algorithm after the interim analysis: Analysis for the primary research question – confidence intervals

If analysis does not adjust for the possibility of adapting the COBIx algorithm, the naïve 95% confidence interval for a diagnostic measure can be obtained using the expression,

$$\text{Estimate} \pm 1.96\text{SE},$$

where expressions for estimate and standard error (SE) are as given in the last two columns in Table 5. Naïve confidence intervals may not achieve nominal coverage probability and may not be symmetric in their coverage, with probability that the confidence interval is above the true diagnostic measure likely to be above 2.5%. Therefore, adjusted confidence intervals are needed.

Like point estimates, existing confidence intervals that adjust for interim analysis decision do not adjust for correlations between sensitivity, specificity, NPV and PPV. Therefore, building on Pepe *et al.* (2009) and the new point estimators we have developed to analyse this study, we have developed a new parametric bootstrap procedure to compute confidence intervals for this study. The confidence intervals computed using this procedure are compared to existing confidence intervals in the same manuscript that includes new point estimates and are shown to have the best properties. As before, based on diagnostic categories by the pathologists, let n_1^+ and n_1^- respectively denote stage 1 number of abnormal and normal biopsies. Similarly, based on diagnostic categories by the pathologists, let n_2^+ and n_2^- respectively denote stage 2 number of abnormal and normal biopsies. The bootstrap algorithm procedure will include the following steps

i. Noting the UMVCUEs (the point estimates) using the COBIx study data that are computed as described above.
ii. Simulating a single study bootstrap data: Using UMVCUEs for sensitivity ($\hat{\theta}_{S,UMVCUE}$) and specificity ($\hat{\theta}_{Sp,UMVCUE}$) in step (i), simulate stage 1 data where TP and TN are simulated assuming $\text{Binomial}\left(n_1^+, \hat{\theta}_{S,UMVCUE}\right)$ and $\text{Binomial}\left(n_1^-, \hat{\theta}_{Sp,UMVCUE}\right)$. Compute stage 1 estimates for sensitivity, specificity, NPV and PPV. If for the simulated stage 1 data, sensitivity ≥ 0.94 and NPV ≥ 0.88, simulate stage 2 data where TP and TN are simulated assuming $\text{Binomial}\left(n_2^+, \hat{\theta}_{S,UMVCUE}\right)$ and $\text{Binomial}\left(n_2^-, \hat{\theta}_{Sp,UMVCUE}\right)$.
iii. Simulating more studies until 5,000 studies that simulate stage 2: Repeat step (ii) until 5,000 studies that simulate stage 2 data.
iv. For each of the 5,000 simulated studies in step (ii) and step (iii) that simulate stage 2 data, use simulated stage 1 data to compute stage 1 correlations between sensitivity, specificity, NPV and PPV as described in the second paragraph in Section 5.3.1.5. These correlations are used in the next step.
v. For each of the 5,000 simulated studies in step (ii) and step (iii) that simulate stage 2 data, compute UMVCUEs for sensitivity, specificity, NPV and PPV using expressions (3) to (6). For each of the 5,000 studies, the correlations required in expressions (3) and (6) are those obtained in step (iv).
vi. For each diagnostic measure, the lower and upper bounds of the 95% confidence interval are the 2.5% and 97.5% percentiles of its 5,000 estimates computed in step (v).

The primary analysis will be the adjusted confidence intervals. Naïve confidence intervals will be reported for sensitivity analysis. Table 8 gives a template for summarising the results if the COBIx algorithm is not adapted after the interim analysis. The R code that will be used to compute the confidence intervals is provided in Appendix 2.

Table 8: Template table for summarising results if the COBIx algorithm is not modified after the interim analysis

Diagnostic measure	Estimate (95% confidence interval)	
	Primary analysis (Adjusted for interim analysis)	**Sensitivity analysis (Naïve analysis)**
Sensitivity Specificity Negative predictive value (NPV) Positive predictive value (PPV)		

5.3.2 Subgroup analyses

The following characteristics define clinically important subgroups:

- Sex
- Ethnicity: White ethnicity versus non-White ethnicity.
- Age: Using the median age from the observed study data.
- Neoplastic versus non-neoplastic.
- Screening biopsies, such as those from breast cancer screening programmes, versus non-screening biopsies.
- Routine versus rare/unusual entities.

5.4 Missing data

Missing data can only arise if the COBIx algorithm cannot read a biopsy image. There is very small chance of this happening and so only complete case analysis will be performed.

5.5 Harms

COBIx study is not an interventional study. Therefore, there is no harm that can arise because of the study procedures. At interim analysis, FNs and FPs will be reviewed by study pathologists to determine if there are patient safety issues to be addressed. This will be repeated at final analysis.

5.6 Statistical software

All analysis will be performed in R statistical program. The analysis code is provided in Appendix 2.

6 References

1) Bainbridge S, Cake R, Meredith M, Furness P, Gordon B. Testing times to come? an evaluation of pathology capacity across the UK., (Cancer Research UK, 2016)
2) Hanna TP et al. Mortality due to cancer treatment delay: systematic review and meta-analysis. *BMJ*, 2020; **371**, m4087.
3) Graham S, Minhas F, Bilal M, Ali M, Tsang YW, Eastwood M, Wahab N, Jahanifar M, Hero E, Dodd K, Sahota H, Wu S, Lu W, Azam A, Benes K, Nimir M, Hewitt K, Bhalerao A, Robinson A, Eldaly H, Raza SEA, Gopalakrishnan K, Snead D, Rajpoot N. Screening of normal endoscopic large bowel biopsies with interpretable graph learning: a retrospective study. *Gut,* 2023; **72**: 1709-1721.
4) Robertson DS, Prevost AT, Bowden J. Accounting for selection and correlation in the analysis of two-stage genome-wide association studies. *Biostatistics*, 2016; **17**: 634-649.
5) Khan JN, Kimani PK, Glimm E, Stallard N. Adjusting for treatment selection in phase II/III clinical trials with time to event data. *Statistics in Medicine,* 2023; **42**: 146-163.
6) Pepe MS, Feng Z, Longton G, Koopmeiners JS. Conditional estimation of sensitivity and specificity from a phase 2 biomarker study allowing early termination for futility. *Statistics in Medicine,* 2009; **28**: 762-779.

Appendix 1: Simulation study used to choose decision boundaries

Simulations setting

In the simulations, we will assume the following are known. However, we will consider several scenarios by varying their values:

- Disease prevalence. Data from seven sites included in the proposal indicated prevalence of 0.65. Because of enriching in the study, we will also consider prevalences of 0.60 and 0.70.
- Sensitivity. The target sensitivity is 0.95. However, we will consider scenarios with sensitivity in the range 0.94-0.97.
- Specificity. Sensitivity is related to specificity. We will use the relationship in the ROC curve with the data used to develop the first version of the COBI algorithm (Graham *et al.*, 2023). For example, when sensitivity is 0.95, specificity is 0.85.
- Total sample stage 1 size n_1. Analysis will be performed in September 2024 using cases provided by end of August 2024. The target sample size by end of August 2024 is 350 retrospective cases by each site, 3,850 cases by all sites. In simulations, we consider a sample size of 3,800 (after rounding down to a hundred and assuming one biopsy per case), a sample size of 7,600 (assuming two biopsies per case) and a sample size 11,400 (assuming three biopsies per case).

Note that NPV and PPV can be computed when prevalence, sensitivity and specificity are known:

$$NPV = \frac{(1-\text{prevalence}) \times \text{specificity}}{[\text{prevalence} \times (1-\text{sensitivity})] + [(1-\text{prevalence}) \times \text{specificity}]},$$

and

$$PPV = \frac{\text{prevalence} \times \text{sensitivity}}{[\text{prevalence} \times \text{sensitivity}] + [(1-\text{prevalence}) \times (1-\text{specificity})]}.$$

We simulate stage 1 data, which corresponds to Table A1 below, which is a snippet of Table 3. Because disease prevalence is assumed to be known, $A_1 + C_1$ and $B_1 + D_1$ are consequently assumed to be known.

Table A1: Stage 1 data summary

COBI algorithm	**Pathologists' diagnoses (Reference standard)**		
	Disease	**No disease**	**Total**
Disease	A_1 (TP)	B_1 (FP)	$A_1 + B_1$
No disease	C_1 (FN)	D_1 (TN)	$C_1 + D_1$
Total	$A_1 + C_1$	$B_1 + D_1$	$n_1 = A_1 + B_1 + C_1 + D_1$

For each simulated study, we:

i. Draw A_1 (TP) taking $A_1 \sim Binomial(A_1 + C_1, \text{sensitivity})$.
ii. Draw D_1 (TN) taking $D_1 \sim Binomial(C_1 + D_1, \text{specificity})$.

iii. Compute sensitivity and NPV.
iv. For several values for F_1 (boundary for sensitivity) in the range 0.94 to 0.97and F_2 (boundary for NPV) in the range 0.88 to 0.90, check if sensitivity ≥ F_1 and NPV ≥ F_2.

For each scenario, 1,000,000 studies were simulated. For each pair of F_1 and F_2, the simulated probability of not adapting the COBIx algorithm is the proportion of simulated studies (out of 1,000,000) with sensitivity ≥ F_1 and NPV ≥ F_2.

Description of simulation results

Simulation results when prevalence is 0.65 are summarised in Figure A1. From top to bottom, different rows correspond to stage 1 sample sizes 3800, 7600 and 11400, respectively. From left to right, different columns correspond to NPV boundaries (F_2) of 0.88, 0.89 and 0.90, respectively. The black, red, green and blue lines correspond to sensitivity-specificity pairs (0.97, 0.725), (0.96, 0.800) (0.95, 0.850) and (0.94, 0.880), respectively. The sensitivity-specificity pairs were extracted from the ROC curve in Graham *et al.*, 2023. NPV and PPV depend on prevalence (assumed to be 0.65) and are computed using expressions given above. The green lines correspond to a scenario for which the ideal decision is to adapt the COBIx algorithm because sensitivity is below 0.95. For the other scenarios, ideal decision is not to adapt the COBIx algorithm.

For each plot in Figure A1, y-axis gives the probability of continuing to stage 2 and x-axis gives the four possible futility boundaries (0.94, 0.95, 0.96 and 0.97).

Choosing decision boundaries

With decision boundaries $F_1 = 0.94$ and $F_2 = 0.88$ for sensitivity and specificity, respectively, the probability of not adapting COBIx algorithm when the ideal decision is not to adapt (scenarios corresponding to green, red and black in Figure A1) is high. Other boundaries may sometimes lead to substantial probability of adapting COBIx algorithm (as high as 0.5 in Figure A1). The results are similar for other values of prevalence (Figure A2 and Figure A3). Therefore, the study management group (SMG) decided to set decision boundaries for sensitivity and NPV as $F_1 = 0.94$ and $F_2 = 0.88$, respectively.

We note that with the chosen boundaries, there is potential of high probability of not adapting COBIx algorithm for scenarios where the ideal decision is to adapt COBIx algorithm. An example is the scenario corresponding to the blue lines in Figure A1, Figure A2 and Figure A3 where probabilities of not adapting is around 0.50 for the different sample sizes and prevalences.

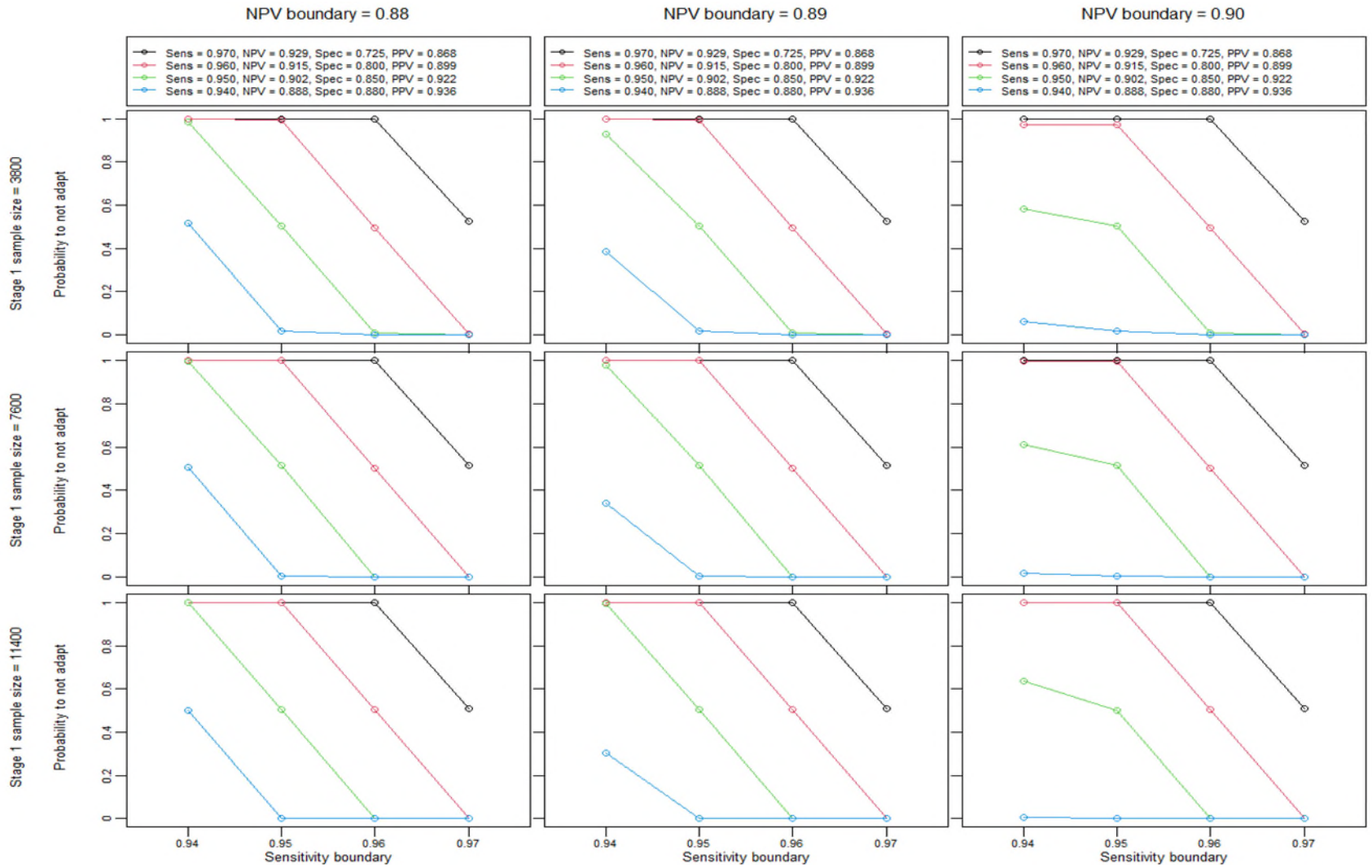

Figure A1: Probability of not adapting the COBIx algorithm after interim analysis for various scenarios. Disease prevalence is 0.65 in all scenarios. Different rows correspond to different stage 1 sample sizes. Different line colours correspond to different combination of sensitivity and specificity.

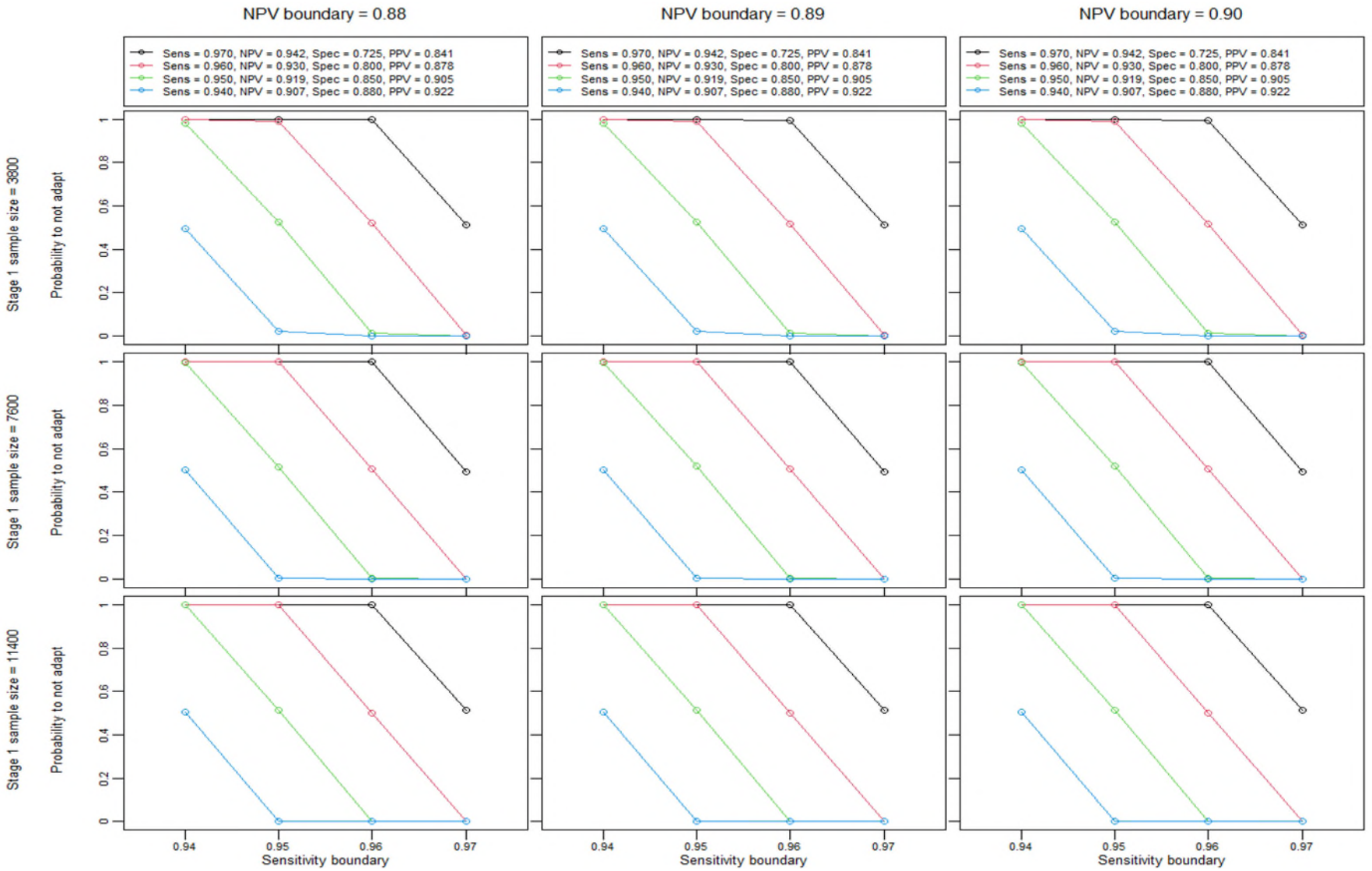

Figure A2: Probability of not adapting the COBIx algorithm after interim analysis for various scenarios. Disease prevalence is 0.60 in all scenarios. Different rows correspond to different stage 1 sample sizes. Different line colours correspond to different combination of sensitivity and specificity.

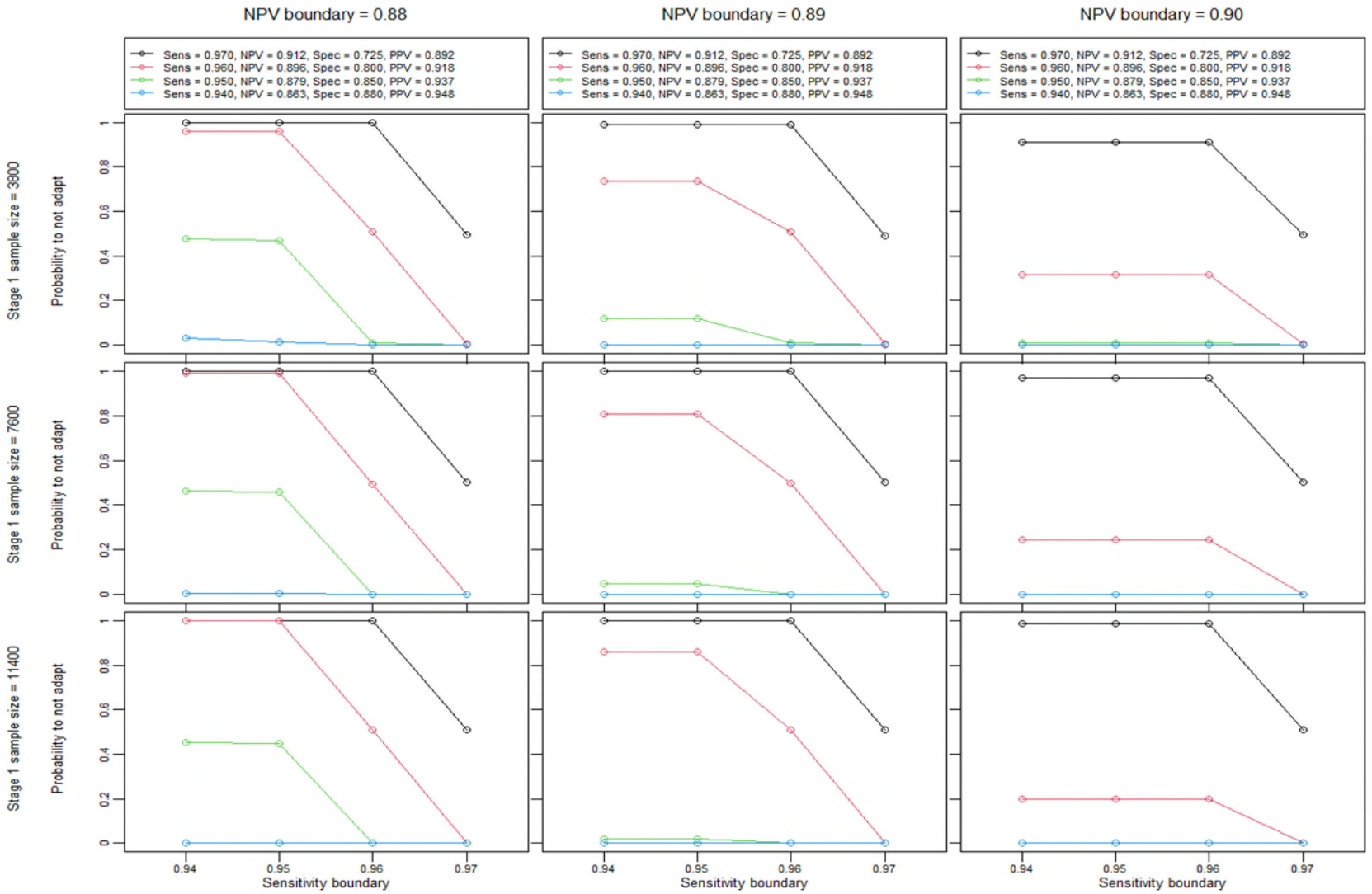

Figure A3: Probability of not adapting the COBIx algorithm after interim analysis for various scenarios. Disease prevalence is 0.70 in all scenarios. Different rows correspond to different stage 1 sample sizes. Different line colours correspond to different combination of sensitivity and specificity.

Appendix 2: R Code for performing primary analysis.

```
##################################
## Interim Analysis Boundaries ##
##################################

boundarySENS <- 0.94 # Sensitivity boundary
boundaryNPV <- 0.88 # NPV boundary


#################################################
########## Stage 1 confusion matrix data ##########
#### They will computed from the study dataset ####
######### Computed using COBIx version 1 ##########
#################################################

tp1 <- 1900 # Number of stage 1 TPs (An R object from raw data computation to be assigned)
fn1 <- 100 # Number of stage 1 FNs (An R object from raw data computation to be assigned)

tn1 <- 900 # Number of stage 1 TNs (An R object from raw data computation to be assigned)
fp1 <- 100 # Number of stage 1 FPs (An R object from raw data computation to be assigned)


#################################################
########## Stage 2 confusion matrix data ##########
#### They will computed from the study dataset ####
#################################################

tp2 <- 1890 # Number of stage 2 TPs (An R object from raw data computation to be assigned)
fn2 <- 110 # Number of stage 2 FNs (An R object from raw data computation to be assigned)
```

```
tn2 <- 900 # Number of stage 2 TNs (An R object from raw data computation to be assigned)
fp2 <- 100 # Number of stage 2 FPs (An R object from raw data computation to be assigned)


#####################################################################
##### CODE FOR COMPUTING DIAGNOSTIC MEASURES AND 95% CIs ######
#####################################################################

SimpleDiagnosticAnalysis <- function(tp, fn, tn, fp){
        PointEstSensitivity <- tp/(tp + fn)
        SeSensitivity <- sqrt((PointEstSensitivity * (1 - PointEstSensitivity))/(tp + fn))
        LbSensitivity <- PointEstSensitivity - (qnorm(0.975) * SeSensitivity)
        UbSensitivity <- PointEstSensitivity + (qnorm(0.975) * SeSensitivity)
        PointEstSensitivity <- format(round(PointEstSensitivity, 3), nsmall=3)
        LbSensitivity <- format(round(LbSensitivity, 3), nsmall=3)
        UbSensitivity <- format(round(UbSensitivity, 3), nsmall=3)
        SensitivityResults <- paste(PointEstSensitivity, " (", LbSensitivity, ", ", UbSensitivity, ")", sep="")

        PointEstSpecificity <- tn/(tn + fp)
        SeSpecificity <- sqrt((PointEstSpecificity * (1 - PointEstSpecificity))/(tn + fp))
        LbSpecificity <- PointEstSpecificity - (qnorm(0.975) * SeSpecificity)
        UbSpecificity <- PointEstSpecificity + (qnorm(0.975) * SeSpecificity)
        PointEstSpecificity <- format(round(PointEstSpecificity, 3), nsmall=3)
        LbSpecificity <- format(round(LbSpecificity, 3), nsmall=3)
        UbSpecificity <- format(round(UbSpecificity, 3), nsmall=3)
        SpecificityResults <- paste(PointEstSpecificity, " (", LbSpecificity, ", ", UbSpecificity, ")", sep="")

        PointEstPpv <- tp/(tp + fp)
        SePpv <- sqrt((PointEstPpv * (1 - PointEstPpv))/(tp + fp))
        LbPpv <- PointEstPpv - (qnorm(0.975) * SePpv)
        UbPpv <- PointEstPpv + (qnorm(0.975) * SePpv)
        PointEstPpv <- format(round(PointEstPpv, 3), nsmall=3)
```

```
        LbPpv <- format(round(LbPpv, 3), nsmall=3)
        UbPpv <- format(round(UbPpv, 3), nsmall=3)
        PpvResults <- paste(PointEstPpv, " (", LbPpv, ", ", UbPpv, ")", sep="")

        PointEstNpv <- tn/(tn + fn)
        SeNpv <- sqrt((PointEstNpv * (1 - PointEstNpv))/(tn + fn))
        LbNpv <- PointEstNpv - (qnorm(0.975) * SeNpv)
        UbNpv <- PointEstNpv + (qnorm(0.975) * SeNpv)
        PointEstNpv <- format(round(PointEstNpv, 3), nsmall=3)
        LbNpv <- format(round(LbNpv, 3), nsmall=3)
        UbNpv <- format(round(UbNpv, 3), nsmall=3)
        NpvResults <- paste(PointEstNpv, " (", LbNpv, ", ", UbNpv, ")", sep="")

        DiagnosticResultsSummary <- list(Sensitivity=SensitivityResults, Specificity=SpecificityResults, Ppv=PpvResults, Npv=NpvResults)
        return(DiagnosticResultsSummary)
}



################################################################################################
################################################################################################
############ CASE OF MODIFYING THE COBIx ALGORITHM AFTER THE INTERIM ANALYSIS ###############
################################################################################################
################################################################################################

############ PRIMARY ANALYSIS - USES STAGE 2 DATA ONLY #############

PrimaryAnalysisResult <- SimpleDiagnosticAnalysis(tp=tp2, fn=fn2, tn=tn2, fp=fp2) # Computing diagnostic measures (95% CI) using stage 2 data only

PrimaryAnalysisResult$Sensitivity
PrimaryAnalysisResult$Specificity
PrimaryAnalysisResult$Npv
```

```
##################################################################################################
############## CASE OF NOT MODIFYING THE COBIx ALGORITHM AFTER THE INTERIM ANALYSIS ###############
##################################################################################################

#####################################################################
##### FUNCTION FOR COMPUTING ADJUSTED DIAGNOSTIC MEASURES ######
#####################################################################

UMVCUEScomputation <- function(stage1tp, stage1fn, stage1tn, stage1fp, stage2tp, stage2fn, stage2tn, stage2fp, sensboundary, npvboundary, nBoots){

        tp <- stage1tp
        fn <- stage1fn
        tn <- stage1tn
        fp <- stage1fp

        sens1 <- tp/(tp + fn)
        spec1 <- tn/(tn+fp)
        ppv1 <- tp/(tp+fp)
        npv1 <- tn/(tn + fn)

        cases <- tp + fn
        negatives <- fp + tn

        sigma1.sens <- (tp * fn)/((tp + fn)^3)
        sigma1.spec <- (tn*fp)/((fp + tn)^3)
```

```
sigma1.ppv <- (tp*fp)/((tp+fp)^3)
sigma1.npv <- (tn*fn)/((tn+fn)^3)

tp <- stage2tp
fn <- stage2fn
tn <- stage2tn
fp <- stage2fp

sens2 <- tp/(tp + fn)
spec2 <- tn/(tn + fp)
ppv2 <- tp/(tp + fp)
npv2 <- tn/(tn + fn)

sigma2.sens <- (tp * fn)/((tp + fn)^3)
sigma2.spec <- (tn*fp)/((tn + fp)^3)
sigma2.ppv <- (tp*fp)/((tp+fp)^3)
sigma2.npv <- (tn*fn)/((tn+fn)^3)


tp <- stage1tp + stage2tp
fn <- stage1fn + stage2fn
tn <- stage1tn + stage2tn
fp <- stage1fp + stage2fp

sensCombined <- tp/(tp + fn)
specCombined <- tn/(tn + fp)
ppvCombined <- tp/(tp + fp)
npvCombined <- tn/(tn + fn)


tp <- rbinom(nBoots, cases, sens2)
fn <- cases - tp
```

```
        tn <- rbinom(nBoots, negatives, spec2)
        fp <- negatives - tn

        sensRho <- tp/(tp + fn)
        npvRho <- tn/(tn + fn)
        specRho <- tn/(tn+fp)
        ppvRho <- tp/(tp+fp)

        RhoSensNPV <- cor(sensRho, npvRho)
        RhoSensNPV <- max(0, RhoSensNPV)
        RhoSensPPV <- cor(sensRho, ppvRho)
        RhoSensPPV <- max(0, RhoSensPPV)
        RhoSpecNPV <- cor(specRho, npvRho)
        RhoSpecNPV <- max(0, RhoSpecNPV)
        RhoSpecPPV <- cor(specRho, ppvRho)
        RhoSpecPPV <- max(0, RhoSpecPPV)
        RhoNpvPpv <- cor(ppvRho, npvRho)
        RhoNpvPpv <- max(0, RhoNpvPpv)

rm(sensRho, npvRho, specRho, ppvRho)


        # Sensitivity

        ThetaSensSens <- sens1 + ((sigma1.sens/sigma2.sens) * sens2)
        ThetaNpvSens <- npv1 + (((RhoSensNPV*sqrt(sigma1.npv)*sqrt(sigma1.sens))/sigma2.sens) * sens2)

        U1S <- (sigma2.sens/sigma1.sens) * (ThetaSensSens - sensboundary)
        U2S <- (sigma2.sens/(RhoSensNPV*sqrt(sigma1.npv)*sqrt(sigma1.sens))) * (ThetaNpvSens - npvboundary)
if (RhoSensNPV > 0){US <- min(U1S, U2S)}
if (RhoSensNPV == 0){US <- U1S}
```

```
        WtSens <- ((sens1 * sigma2.sens)+(sens2*sigma1.sens))/(sigma1.sens+sigma2.sens)
        sigma1 <- sigma2.sens/sqrt(sigma1.sens + sigma2.sens)
        sigma2 <- (sqrt(sigma2.sens + sigma1.sens))/sigma1.sens
        WSU <- sigma2 * (US - WtSens)
        AdjSens <- sensCombined - (sigma1 * ((dnorm(WSU))/(pnorm(WSU))))


        # Specificity

if (RhoSpecNPV > 0){
        ThetaNpvSpec <- npv1 + (((RhoSpecNPV*sqrt(sigma1.npv)*sqrt(sigma1.spec))/sigma2.spec) * spec2)
        USp <- (sigma2.spec/(RhoSpecNPV*sqrt(sigma1.npv)*sqrt(sigma1.spec))) * (ThetaNpvSpec - npvboundary)

        WtSpec <- ((spec1 * sigma2.spec)+(spec2*sigma1.spec))/(sigma1.spec+sigma2.spec)
        sigma1 <- sigma2.spec/sqrt(sigma1.spec + sigma2.spec)
        sigma2 <- (sqrt(sigma2.spec + sigma1.spec))/sigma1.spec
        WSpU <- sigma2 * (USp - WtSpec)
        AdjSpec <- specCombined - (sigma1 * ((dnorm(WSpU))/(pnorm(WSpU))))
}
if (RhoSpecNPV == 0){AdjSpec <- specCombined}

        # NPV

        ThetaSensNpv <- sens1 + (((RhoSensNPV*sqrt(sigma1.sens)*sqrt(sigma1.npv))/sigma2.npv) * npv2)
        ThetaNpvNpv <- npv1 + ((sigma1.npv/sigma2.npv) * npv2)

        U1N <- (sigma2.npv/(RhoSensNPV*sqrt(sigma1.sens)*sqrt(sigma1.npv))) * (ThetaSensNpv - sensboundary)
        U2N <- (sigma2.npv/sigma1.npv) * (ThetaNpvNpv - npvboundary)
if (RhoSensNPV > 0){UN <- min(U1N, U2N)}
if (RhoSensNPV == 0){UN <- U2N}

        WtNpv <- ((npv1 * sigma2.npv)+(npv2*sigma1.npv))/(sigma1.npv+sigma2.npv)
```

```
        sigma1 <- sigma2.npv/sqrt(sigma1.npv + sigma2.npv)
        sigma2 <- (sqrt(sigma2.npv + sigma1.npv))/sigma1.npv
        WNU <- sigma2 * (UN - WtNpv)
        AdjNpv <- npvCombined - (sigma1 * ((dnorm(WNU))/(pnorm(WNU))))


        # PPV

        ThetaSensPpv <- sens1 + (((RhoSensPPV*sqrt(sigma1.sens)*sqrt(sigma1.ppv))/sigma2.ppv) * ppv2)
        ThetaNpvPpv <- npv1 + (((RhoNpvPpv*sqrt(sigma1.npv)*sqrt(sigma1.ppv))/sigma2.ppv) * ppv2)

        U1P <- (sigma2.ppv/(RhoSensPPV*sqrt(sigma1.sens)*sqrt(sigma1.ppv))) * (ThetaSensPpv - sensboundary)
        U2P <- (sigma2.ppv/(RhoNpvPpv*sqrt(sigma1.npv)*sqrt(sigma1.ppv))) * (ThetaNpvPpv - npvboundary)
        CorIndic <- ((RhoSensPPV > 0) * 1) + ((RhoNpvPpv > 0) * 10)
if (CorIndic == 11){UP <- min(U1P, U2P)}
if (CorIndic == 1){UP <- U1P}
if (CorIndic == 10){UP <- U2P}

        WtPpv <- ((ppv1 * sigma2.ppv)+(ppv2*sigma1.ppv))/(sigma1.ppv+sigma2.ppv)
        sigma1 <- sigma2.ppv/sqrt(sigma1.ppv + sigma2.ppv)
        sigma2 <- (sqrt(sigma2.ppv + sigma1.ppv))/sigma1.ppv
        WPU <- sigma2 * (UP - WtPpv)
        AdjPpv <- ppvCombined - (sigma1 * ((dnorm(WPU))/(pnorm(WPU))))
if (CorIndic == 0){AdjPpv <- ppvCombined}

        AdjustedEstimates <- list(AdjSensitivity = AdjSens, AdjSpecificity = AdjSpec, AdjPPV = AdjPpv, AdjNPV = AdjNpv)
        return(AdjustedEstimates)

}


############################################################
```

```
#################### COMPUTING UMVCUES ####################
############################################################

SimulationSeed <- 67435 # Seed

set.seed(SimulationSeed) # Setting seed for reproducibility
UMVCUEestimates <- UMVCUEScomputation(stage1tp=tp1, stage1fn=fn1, stage1tn=tn1, stage1fp=fp1, stage2tp=tp2, stage2fn=fn2, stage2tn=tn2, stage2fp=fp2,
sensboundary=boundarySENS, npvboundary=boundaryNPV, nBoots=5000)
#UMVCUEestimates


#####################################################################
##### FUNCTION FOR COMPUTING ADJUSTED CONFIDENCE INTERVALS ######
#####################################################################

BootClsX <- function(nsimConf, BootsN, cases, negatives, cases2, negatives2, sens, spec, BoundaryVector){
        tp <- rbinom(nsimConf, cases, sens)
        fn <- cases - tp
        tn <- rbinom(nsimConf, negatives, spec)
        fp <- negatives - tn

        BootstrapMatrix <- matrix(NA, nrow=nsimConf, ncol=8)

        BootstrapMatrix[,1] <- seq(1, nsimConf)
        BootstrapMatrix[,3:6] <- cbind(tp, fp, fn, tn)

        sens1 <- tp/(tp + fn)
        npv1 <- tn/(tn + fn)
        StopBoots <- 1 * (sens1 >= BoundaryVector[1]) * (npv1 >= BoundaryVector[2])
        BootstrapMatrix[,8] <- StopBoots

        NumSims <- cumsum(StopBoots)
```

```
        NumSims1 <- 1 * (NumSims < BootsN)
        NumSims2 <- min(length(BootstrapMatrix[,8]), sum(NumSims1) + 1)
        BootstrapMatrix <- BootstrapMatrix[1:NumSims2, ]

        ContinueIndicsBoots <- BootstrapMatrix[,1] * BootstrapMatrix[,8]
        ContinueIndicsBoots <- setdiff(ContinueIndicsBoots, 0)
        nsimConf1 <- length(ContinueIndicsBoots)
        BootstrapMatrix <- BootstrapMatrix[ContinueIndicsBoots, ]
        tp1 <- BootstrapMatrix[,3]
        fp1 <- BootstrapMatrix[,4]
        fn1 <- BootstrapMatrix[,5]
        tn1 <- BootstrapMatrix[,6]

rm(tp, fp, fn, tn, StopBoots, ContinueIndicsBoots, NumSims, NumSims1)

        DiagnosticCIs <- matrix(NA, ncol=2, nrow=4)

        if (nsimConf1 == BootsN){
                tp2 <- rbinom(nsimConf1, cases2, sens)
                fn2 <- cases2 - tp2

                tn2 <- rbinom(nsimConf1, negatives2, spec)
                fp2 <- negatives2 - tn2

                BootstrapMatrix <- matrix(NA, nrow=nsimConf1, ncol=4)

                for (i in 1:nsimConf1){
                        UMVCUEestimatesCI <- UMVCUEScomputation(stage1tp=tp1[i], stage1fn=fn1[i], stage1tn=tn1[i], stage1fp=fp1[i], stage2tp=tp2[i],
stage2fn=fn2[i], stage2tn=tn2[i], stage2fp=fp2[i], sensboundary=BoundaryVector[1], npvboundary=BoundaryVector[2], nBoots=BootsN)
                        BootstrapMatrix[i,] <- c(UMVCUEestimatesCI$AdjSensitivity, UMVCUEestimatesCI$AdjSpecificity, UMVCUEestimatesCI$AdjPPV,
UMVCUEestimatesCI$AdjNPV)
                }
```

```
        DiagnosticCIs[1,] <- quantile(BootstrapMatrix[,1], c(0.025, 0.975))
        DiagnosticCIs[2,] <- quantile(BootstrapMatrix[,2], c(0.025, 0.975))
        DiagnosticCIs[3,] <- quantile(BootstrapMatrix[,3], c(0.025, 0.975))
        DiagnosticCIs[4,] <- quantile(BootstrapMatrix[,4], c(0.025, 0.975))
        }
        AdjustedCIs <- list(AdjSensitivityCIs = DiagnosticCIs[1,], AdjSpecificityCIs = DiagnosticCIs[2,], AdjPPVCIs = DiagnosticCIs[3,], AdjNPVCIs = DiagnosticCIs[4,])
        return(AdjustedCIs)
}


#################################################################
################## COMPUTING ADJUSTED CIs ##################
#################################################################

stage1cases <- tp1 + fn1  ## Stage 1 number of abnormal biopsies
stage1negatives <- fp1 + tn1  ## Stage 1 number of normal biopsies
stage2cases <- tp2 + fn2  ## Stage 2 number of abnormal biopsies
stage2negatives <- fp2 + tn2  ## Stage 2 number of normal biopsies


SimulationSeed <- 6274359 # Seed
set.seed(SimulationSeed) # Setting seed for reproducibility
ConfInts <- BootCIsX(nsimConf=1000000, BootsN=5000, cases=stage1cases, negatives=stage1negatives, cases2=stage2cases, negatives2=stage2negatives,
sens=UMVCUEestimates$AdjSensitivity, spec=UMVCUEestimates$AdjSpecificity, BoundaryVector=c(boundarySENS, boundaryNPV))
#ConfInts


#############################################################################
############## PRINTING ALL (ESTIMATES AND 95% CIs) ADJUSTED RESULTS ##############
#############################################################################

a <- format(round(UMVCUEestimates$AdjSensitivity, 3), nsmall=3)
```

```
b <- format(round(ConfInts$AdjSensitivityCIs[1], 3), nsmall=3)
c <- format(round(ConfInts$AdjSensitivityCIs[2], 3), nsmall=3)
AdjustedAnalysisSensitivity <- paste(a, " (", b, ", ", c, ")", sep="")

a <- format(round(UMVCUEestimates$AdjSpecificity, 3), nsmall=3)
b <- format(round(ConfInts$AdjSpecificityCIs[1], 3), nsmall=3)
c <- format(round(ConfInts$AdjSpecificityCIs[2], 3), nsmall=3)
AdjustedAnalysisSpecificity <- paste(a, " (", b, ", ", c, ")", sep="")

a <- format(round(UMVCUEestimates$AdjPPV, 3), nsmall=3)
b <- format(round(ConfInts$AdjPPVCIs[1], 3), nsmall=3)
c <- format(round(ConfInts$AdjPPVCIs[2], 3), nsmall=3)
AdjustedAnalysisPPV <- paste(a, " (", b, ", ", c, ")", sep="")

a <- format(round(UMVCUEestimates$AdjNPV, 3), nsmall=3)
b <- format(round(ConfInts$AdjNPVCIs[1], 3), nsmall=3)
c <- format(round(ConfInts$AdjNPVCIs[2], 3), nsmall=3)
AdjustedAnalysisNPV <- paste(a, " (", b, ", ", c, ")", sep="")

AdjustedAnalysisSensitivity
AdjustedAnalysisSpecificity
AdjustedAnalysisNPV
AdjustedAnalysisPPV



################################################################################
############ SENSITIVITY ANALYSIS - NAIVE POINT ESTIMATES AND 95% CIs #############
################################################################################

NaiveAnalysisResult <- SimpleDiagnosticAnalysis(tp=tp1+tp2, fn=fn1+fn2, tn=tn1+tn2, fp=fp1+fp2)
NaiveAnalysisResult$Sensitivity
```

NaiveAnalysisResult$Specificity
NaiveAnalysisResult$Npv
NaiveAnalysisResult$Ppv