FOCUS4

Molecular selection of therapy in colorectal cancer: a molecularly-stratified randomised controlled trial programme.

MRC CTU ID: CR13

ISRCTN #: ISRCTN#90061546

EUDRACT #: 2012-005111-12

CTA #: 00316/0245/001-0001

REC #: 13/SC/0111

STATISTICAL ANALYSIS PLAN

Version 3.0

MRC Clinical Trials Unit

Version date: 2nd November 2020

Contacts:
Programme Lead: Rick Kaplan
Project Lead: Louise Brown
Statistician: David Fisher

Tel: +44 (0)20 7670 4653
Fax: +44 (0)20 7670 4818
Email: mrcctu.focus4@ucl.ac.uk

This template and all preceding versions will be stored in the Statistical Analysis Master File for this trial held by David Fisher.

Contents

Important

This is a generic Statistical Analysis Plan, relevant to each Trial in the FOCUS4 Trials Programme. The term "Trial" is to be interpreted in this document as "a member of the FOCUS4 Trials Programme". Where a specific Trial deviates from this Plan, the details will be made clear either in this Plan or in a File Note.

1. Background and Design

FOCUS4 is a molecularly stratified, multi-arm, multi-stage (MAMS), multi-site randomised trial programme for patients with advanced colorectal cancer. During the initial registration period, all patients are treated with standard chemotherapy and considered for a standard treatment break if they have responding or stable disease after 16 weeks of chemotherapy. During the registration period, biomarker testing will be performed on their original tumour specimens to determine which specific agent(s) may be most appropriate to test during interruption of chemotherapy after 16 weeks. The patient will then be offered entry into a specific Trial on the basis of their molecular cohort. Each of these Trials (which are each identified by a unique letter) will be double blind and placebo controlled for oral agents but may be modified for intravenously administered agents as a double blind placebo design may not be appropriate or acceptable to patients. A separate, specific, Trial Protocol will describe the procedures for that Trial. As new agents are tested within each molecular cohort of the FOCUS4 Trial Programme, a new letter is assigned to that Trial.

Each Trial utilises the Multi-Arm Multi-Stage (MAMS) trial design with staged intermediate analyses reviewed by the Independent Data Monitoring Committee (IDMC). The first two of these analyses will be equivalent to a conventional Phase 2 study to assess safety (Stage I) and lack-of-sufficient-activity (Stage II). At this point, results from Stages I and II may be released outside the IDMC and on the basis of the findings, the Trial will either stop accrual or progress to continued recruitment to assess efficacy for progression-free survival (PFS) (Stage III) and, possibly, efficacy for overall survival (OS) (Stage IV) a potential additional primary outcome. Continuation to further accrual in these additional stages, which will be equivalent to a conventional Phase 3 study, will depend on the strength of effect (MAMS-defined critical hazard ratios) seen at the end of Stages I and II and the availability of resources to achieve adequate recruitment and follow-up, including the necessary commitment of supply of the novel agent(s). Further details can be found in the FOCUS4 Master Protocol.

2. Sample Size Calculations

The COIN trial recruited 2,445 patients over 38 months, four months ahead of schedule: an average rate of 60 patients per month. FOCUS 3 tested the feasibility of recruiting patients according to their biomarker panel classification and also managed to complete accrual on schedule by April 2011. For FOCUS4, we have assumed 70 patients will be screened for registration per month when all sites are open. Of these, it is anticipated that 32 will be eligible and consent to randomisation across all molecular cohorts.

Reasons for eligibility for randomisation are as follows:

- 72% are expected to have normal platelets at registration. Those with abnormal platelets at registration will not be registered
  → 50 patients per month registered;
- 90% of these will have stable or responding disease by their interim CT scan
  → 45 patients per month;
- 80% of these should have stable or responding disease by 16 weeks
  → 36 patients per month;
- 88% of these are likely to accept randomisation
  → 32 patients per month randomised.

Next, we estimate the proportion of patients expected to fall into each of the four molecular cohorts to be as follows (with the corresponding trial identifier in brackets):

- 8% BRAF mutated (Trial A)
- 30% PIK3CA mutation or PTEN loss (Trial B)
- 33% KRAS/NRAS mutation (Trial C)
- 27% All wild type for above mutations (Trial D)
- 2% Unclassified (Trial N)

Sample size calculations were carried out based on the above assumptions with the `nstage` program in Stata versions 12.1 to 16.1, which uses a MAMS design incorporating multiple interim analyses for safety, lack-of-sufficient-activity (LSA) and efficacy. This allows non-beneficial comparisons to be identified and halted as soon as possible, with minimal risk of prematurely stopping beneficial comparisons by chance. To achieve this, the alpha value is set initially high (one-sided α=0.30) and is thereafter progressively lowered such that the final efficacy analyses use values of a similar magnitude to conventional statistical tests. Within each biomarker-defined Trial (for each active agent vs. placebo comparison) there are four analysis stages: safety (Stage I), lack-of-sufficient-activity (Stage II), efficacy for PFS (Stage III) and efficacy for OS (Stage IV). Interim results from each stage will be reviewed by the IDMC to guide their recommendations for early termination or continuation of a Trial. In addition, results may be released publicly at the end of Stage II (equivalent to a phase 2 study). Thus, the rationale for either recommending termination or

continuation of the Trial will be transparent to patients, clinicians and providers of the novel agent under scrutiny.

For each Trial, the overall power is maintained at 80%, allowing for multiple interim analyses, with a maximum 5% two-sided overall significance level and a default allocation ratio of 2:1 in favour of the active arm. This ratio has been selected because it provides more information on early safety and toxicity in the active arm. A generic calculation may be found in the Master Protocol; specific calculations for other Trials may be found in the relevant Trial Protocols.

3. Randomisation method

Within each Trial, patients are randomised either to a placebo or to a new targeted agent/combination specific to their biomarker cohort. The specific allocation ratio will be defined and justified within each Trial Protocol. Where possible, to maximise information on novel agents, a 2:1 allocation ratio in favour of the novel agents will be used. Randomisation will be performed using the method of minimisation with a random element. There is an 80% chance of getting the 'minimum' treatment and a 20% chance of treatment being allocated randomly. Minimisation will be stratified by a number of factors known to be prognostic of outcome, as well as the regime used during the 16 weeks of first-line chemotherapy. The global list of minimisation factors, applicable to all Trials, is as follows:

- Randomising hospital site
- Site of primary tumour (Right colon; Left colon; Rectum)
- WHO Performance Status (0/1/2)
- 16-week CT scan result (Stable disease; Partial response; Complete response)
- Number of metastatic sites (0/1/2+)
- First-line chemotherapy regimen
    - Fluoropyrimidine (5FU; Capecitabine; Neither)
    - Oxaliplatin/irinotecan (Both; Ox only; Ir only; Neither)
    - Monoclonal antibody (Cetuximab/Panitumumab; Bevacizumab; None)

In addition, in FOCUS4-N randomisation is also minimised on biomarker cohort, of which the precise categories have changed during the lifetime of the FOCUS4 trial as follows:

- BRAF mut; PIK3CA mut/PTEN loss; KRAS/NRAS mut; All wild type; non-stratified (March 2014 to April 2016)
- BRAF mut; KRAS/NRAS mut; All wild type; PIK3CA mut; non-stratified (April 2016 to August 2017)
- BRAF mut; All wild type; PIK3CA mut; RAS+p53 mut; non-stratified (August 2017 to July 2018)
- BRAF mut; All wild type; PIK3CA mut; RAS+p53 mut; PIK3CA+RAS+p53 mut; non-stratified (July 2018 to March 2020)

4. Data

4.1. CRFs and variables

Full details of data collection and timing are described in the FOCUS4 Master Protocol (version 7.0, 11th September 2019). A copy of the CRFs and Quality of Life (QL) questionnaires are presented in the Trial Master File. Details of the variables are presented as metadata in the Trial Master File.

4.2. Management of datasets

The day-to-day management of the datasets is the responsibility of the trial managers and data managers, but when an analysis is required it is important that there is good communication between the trial/data managers and the trial statistician.

The FOCUS4 Randomisation Database will be locked for the final or interim analyses in accordance with the CTU Database Lock SOP. This includes following procedures for:

- Agreeing in advance levels of quality and completeness of the dataset
- Ensuring that the agreed minimum levels of data quality and completeness have been met
- Requesting a database lock by the trial statistician
- Preparing the database lock document by the Data management Systems Project Manager (DSPM) on consultation with the trial team
- Performing the database lock once the database lock document has been agreed and the SAP has been signed off
- When the database lock has been completed, the database lock document will be signed by the DSPM and trial statistician.
- Un-locking and re-locking the database for any data update requirements

At the time of each interim analysis:

- The trial statistician will file out from Macro a dataset of all data stored in the database. This will act as the frozen dataset. It is the responsibility of the statistician to accurately record the date of freezing and ensure all data is retrieved according to the CTU Database Lock SOP.
- After data extraction, new data can continue to be entered onto the Macro database.
- If any outstanding data queries are resolved during the analysis that relate to data in the frozen dataset (e.g. problems that are found during analysis or amended CRFs that are returned to CTU), the main Macro database should be changed under the oversight of the trial manager. Identical amendments should be made by the statistician in the frozen dataset.

4.3. Data verification

It is important that the data are cleaned to an acceptable level prior to any analysis as erroneous or incomplete data can influence results. Data verification, consistency and range checks will have been performed by the MRC CTU trial and data managers as laid down in the Monitoring, Quality and Data Management Plans, including range, consistency and missing data checks. If appropriate, additional checks will be performed by the trial statistician when the analysis is performed and any variables that demonstrate unusual, outlying, unlabelled or inconsistent values will be queried with the trial/data managers. If possible, data queries will be resolved, although it is accepted that due to administrative reasons and data availability a small number of problems will continue to exist. This will be minimised.

Given the thorough nature of our follow-up procedure we expect the issue of missing data to be relatively minimal. We anticipate high compliance with initial data collection as this is close to the time of patient registration. Therefore, we do not anticipate the use of additional methodology such as imputation.

5. Statistical analysis

5.1. Registration and randomisation

Patients are registered into the FOCUS4 trial programme at the time of commencing first-line chemotherapy (or up to 12 weeks into first-line chemotherapy if otherwise eligible). Registered patients are expected to receive 16 weeks of first-line chemotherapy in total, at the end of which they should have stable or responding disease as measured by an end-of-registration CT scan, and have biomarker results available. If eligible, they are then consented for randomisation. The following data will be presented from the registration period:

- Recruitment
    - A summary of the number of sites that have been opened.
    - Monthly total registrations and randomisations will be plotted in a bar chart along with expected numbers
    - One CONSORT diagram will be updated for patient flow from screening and registration through to randomisation and separate CONSORT diagrams will be provided from randomisation for patient flow through each Trial being analysed.

- Timings, measured in weeks as median and lower/upper quartile:
    - from removal of tumour block to availability of biomarker results (i.e. “age” of tumour block)
    - from registration to availability of biomarker results
    - from registration to randomisation
    - from start to end of first-line chemotherapy
    - from end of first-line chemotherapy to randomisation

    - from 16-week CT scan to randomisation
    - from randomisation to start of first FOCUS4 treatment cycle

- Tabulations of results of CT scans according to RECIST classifications of response
    - At 8 weeks (for those collected)
    - At 16 weeks (for all registered patients who have completed their first-line treatment)

- When ~300-400 biomarker panel results are available, quality assurance (QA) analyses will be carried out, using ANOVA, to check for systematic differences in 16-week CT scan result and biomarker cohort in terms of:
    - Age of tumour block
    - Tumour samples originating from resections vs from biopsies

5.2. Biomarker panel results for registered patients

The proportions expected to fall into each molecular cohort are provided in Section 2 and are estimates from previous COIN trial data. At each interim analysis the observed biomarker panel results and molecular cohort proportions will be tabulated, as follows:

- Numbers and proportions of assay failures
- Tabulation of BRAF status (wild type; mutation), with breakdown of specific mutations
- Tabulation of PIK3CA/PIK3R1 status (wild type; mutation), with breakdown of specific mutations
- Tabulation of PTEN loss (loss vs no loss)
- Tabulation of KRAS status (wild type; mutation), with breakdown of specific mutations
- Tabulation of NRAS status (wild type; mutation), with breakdown of specific mutations
- Tabulation of P53 status after the activation of FOCUS4-C in 2017 (Note that P53 data from before this time is being tested retrospectively for patients randomised before activation of FOCUS4-C. Some samples may not be available leading to incomplete ascertainment of P53 status in all patients)
- MSI status
- Tabulation of molecular cohort eligibility
- Cross-tabulation of molecular cohort eligibility with actual randomisation (including failures to randomise if eligible and reasons for not being randomised)

5.3. Baseline characteristics for randomised patients

To ensure that the minimisation procedure is functioning correctly, baseline characteristics will be presented by randomised group. Categorical variables will be presented as numerators and percentages. Continuous variables will be summarised using the mean and standard deviation, or the median and lower/upper quartile for non-normally distributed variables.

The following baseline characteristics will be presented:

- Sex (categorical)
- Age (continuous, units of years)
- WHO performance status (categorical)
- Disease status by RECIST (categorical)
- Resection status (categorical)
- Number of metastatic sites (categorical 0, 1, 2+)
- Synchronous vs metachronous disease
- First-line therapy:
    - Fluoropyrimidine (categorical)
    - Oxaliplatin/Irinotecan (categorical)
    - Monoclonal antibody (categorical)
- Age of biomarker tissue sample (continuous)
- White blood cell count (continuous, units of 10^9/l)
- Neutrophil count (continuous, units of 10^9/l)
- Platelet count (continuous, units of 10^9/l)
- Serum bilirubin (continuous, units of mmol/l)
- ALP (continuous, units of U/l)
- AST/ALT (continuous, units of U/l)
- Renal function from estimated creatinine clearance or measured GFR (continuous, units of ml/min)
- Calcium (continuous, units of mmol/l)
- Magnesium (continuous, units of mmol/l)
- CEA (continuous, units of µg/l)
- LDH (continuous, units of U/l)
- Use of aspirin/statins/metformin (yes/no)
- EREG expression (continuous, unit-less)

Trial D only:

- ECG result (categorical)
- LVEF result (continuous, units of %)
- Best corrected distance vision (continuous, units of cm, by eye)
- Evidence of maculopathy (yes/no)

Trial C only:

- PIK3CA mutation status (categorical)
- BRAF mutation status (categorical)

Trial N only:

- BRAF mutation status (categorical)
- PIK3CA mutation status (categorical)
- RAS mutation status (categorical)

- P53 mutation status (where available; categorical)
- "All wildtype" status (i.e. any vs none of the above mutations; categorical)
- MSI status (categorical)

5.4. Primary outcome definitions

The primary outcome in FOCUS4 is either progression-free survival or overall survival, depending on the specifics of the MAMS design for each Trial.

Progression-free survival (PFS)

Time from randomisation to first recorded disease progression or death from any cause

Overall survival (OS) (if appropriate – Stage IV only)

Time from randomisation to death from any cause

For both of the above time-to-event outcomes, the time itself will be calculated as the difference of start and stop dates (e.g. date of death minus date of randomisation). By default, observations with zero survival time are excluded by Stata. For this reason, any randomised patients without any post-randomisation follow-up data will be assigned a "time to censoring" of 0.001 days. The number of such patients will be tabulated by arm, with reasons. The number of events observed and median survival and follow-up time will be reported by arm, and survival data presented in a Kaplan-Meier plot with numbers at risk added. Follow-up will also be presented in a Reverse Kaplan-Meier plot.

The following conversion factors will be used to convert days into months or years: 1 month = 30.4375 days, 1 year = 365.25 days.

Censoring criteria

For overall survival, patients will be censored on the date they were last known to be alive, either via collection of prescription from the drug delivery system, the progress form, CT scan date or recorded follow-up appointment date.

Although disease progression according to RECIST v1.1 may be assessed by CT scan or by alternative imaging methods or by clinical assessment alone, progression-free survival *censoring* will only be at the date of a patient's most recent CT scan confirming non-progression.

Patients should have their disease assessed every 8 weeks until progression is confirmed. If a patient dies more than 3 months after their previous CT scan (at which no progression was recorded), we

are unable to assume that no progression has occurred prior to death. Therefore, such patients will be censored at their last CT scan date that confirmed no progression, or at the time of randomisation. For patients who have died prior to any follow-up CT scan, the date of death will be used as the date of the event if it has occurred within *3 months* of randomisation; otherwise the patient will be censored at randomisation.

As a result of the COVID-19 pandemic, recruitment to FOCUS4 ceased on 23rd March 2020 and from that time onwards, patients have been permitted to increase the length of time between CT scans from 8 to a maximum of 12 weeks. Therefore, within the same timeframe, the above criterion should be interpreted with respect to a death observed more than *4 months* from the last CT scan (rather than 3 months).

Patients who are lost to follow-up without any contact after randomisation will be censored and assigned a time of 0.001 days (see above).

Data fields

The following Macro data fields will be used to define the primary outcomes:

Form	Field name	Label	Notes
Patinfo	RandDate	Date of randomisation	Survival origin
Progress4	DateConfirmProg	For PD only, confirm date progression confirmed	For progression-free survival
Progress4	CTscanDate	Date of CT scan assessment	For progression-free survival censoring
Progress1	FollowUpDate	Date of Progress [Form]	For overall survival censoring
Death	DeathDate	Date of death	

5.5. Withdrawals and ineligibility

Withdrawals

If a patient requests to withdraw from the trial, their data will be administratively censored at the date of their withdrawal request. Specific censoring for overall survival and progression-free survival will then be done using the remaining data for that patient. For example, the censoring date for progression-free survival would be the date of the CT scan immediately prior to withdrawal. If a patient requests their data to be removed from the trial entirely, then they will not be counted as having been randomised and will not appear in any counts, tabulations or analyses.

Patient withdrawals are recorded in the TMF, and in an Excel spreadsheet saved on the following subdirectory of the network drive: S:\MRCCTU_Focus4\15. Patients\Withdrawals

Ineligibility

Prior to data freeze, a review of eligibility criteria will be performed to identify patients who may have been incorrectly entered into the trial. All patients deemed to be ineligible will be reviewed by the TMT, and may be excluded from the PPA (but not typically from ITT). Such information will be fully documented, and will need to be hard-coded into the analysis do-files.

Following the intention-to-treat (ITT) principle, the default should be for all randomised patients to be analysed, regardless of retrospective eligibility findings. However, exceptions may be made if there is a strong reason for removal, such as if the cause of the ineligibility could plausibly relate to how their disease might respond to their randomised treatment (e.g. development of resistance to first line therapy during registration; or unreported brain metastases at the time of randomisation). Such exclusions from ITT will be made on a case-by-case basis, and will be fully documented in the TMF with documented approval by the TMT, including a trial clinician and statistician. Such information will also need to be hard-coded into the analysis do-files.

Details of the most recent eligibility review may be found in an Excel spreadsheet saved on the following subdirectory of the network drive: S:\MRCCTU_Focus4\10. Audit and Quality Control Documents\10.3 Internal Audit & Checks\Eligibility Checks\Current Spreadsheets

5.6. Analysis timings

Interim (decision-making) analyses

Analyses will be performed at each interim stage, defined by the number of events accrued in the control arm as described in Appendix 1.

The MAMS design is based on a comparison at each interim stage between the observed (primary adjusted) point estimate of the hazard ratio and a pre-defined critical value, e.g. see Table A2 in Appendix 1. If the observed HR falls above the critical value, this would suggest that there is insufficient evidence of strong enough activity of the treatment and a recommendation to stop the Trial may be made. Conversely, if the observed HR is smaller than the critical value, this provides preliminary evidence that the target final effect size may be observed in the future and the Trial should continue. However, the final decision on whether or not to continue the Trial will be made by the IDMC, who will be presented with all activity, efficacy and toxicity data from the interim analysis.

Final (reporting) analysis

For the final efficacy analysis of a Trial the primary outcome is progression-free survival, unless the IDMC has recommended continuation to Stage IV (overall survival) of the MAMS design. In addition to the Cox proportional-hazards model being fitted, a test of statistical significance will be carried out using a log-rank test, stratified for the minimization factors. A test of the proportional hazards assumption will be performed by regressing scaled Schoenfeld residuals against the log of time (Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. Biometrika 1994; 81:515-26). If there is statistically significant violation of the proportional-hazards assumption, a suitable alternative model will be fitted, such as a Flexible Parametric survival model, or an alternative estimate presented such as restricted mean survival time (which is not dependent on the proportional-hazards assumption), including the same list of covariates as the Cox model.

5.7. Analysis samples

Intention-To-Treat (ITT)

This is the default analysis. According to the intention-to-treat principle, all randomised patients will be included in the analysis and analysed within their randomised group regardless of treatment received. Any patients who are subsequently found to be ineligible at baseline (incorrectly entered into the trial) will also be included in the ITT analysis, unless there is a strong reason to exclude them; see Section 5.5. Patients who withdraw consent for subsequent data collection will also be included (but will be censored; see Section 5.5), but patients who withdraw *all* consent will be removed entirely.

Per-Protocol Analysis (PPA)

The interim analyses for FOCUS4 are being performed in a phase II setting to look for drug activity and some analyses will be performed with very few events (e.g. less than 20). In such cases, a per-protocol analysis will be undertaken as the primary analysis. Under the PPA, patients who stopped active trial treatment (for reasons other than progression), without completing at least one full cycle of treatment, according to the relevant Trial Protocol, will be censored at the time they stopped trial treatment. Patients randomised to an active trial treatment arm who subsequently received no trial treatment at all will be excluded from the PPA. Patients who successfully completed at least one full cycle will have their time to event, or time to censoring, defined as described in Section 5.4. Some patients who are subsequently found to be ineligible at baseline (incorrectly entered into the trial) may also be excluded from the PPA; see Section 5.5.

In particular:

- All patients randomised to Active Monitoring will be included in the PPA unless they receive an anti-cancer treatment prior to progression. They will be censored at the start of their

anti-cancer treatment in the PPA. This will lead to exclusion of some progression events in the AM arm that occur soon after the decision to start anti-cancer treatment without evidence of progression at the time of this decision.

- Patients are regarded as per-protocol so long as the required amount of active treatment was administered within the specified total length of the first cycle, including the time covered by any scheduled breaks of one or more days between doses (or other forms of administration where relevant);

- If the relevant Trial Protocol specifies a maximum cycle delay length, the first cycle must have been started within that length of time from the date of randomisation for a patient to be regarded as per-protocol.

Additionally, patients will be excluded from the PPA if (a) they have been deemed ineligible, e.g. following quality-checking of inclusions using recorded data; AND if (b) following clinical review by Prof. Rick Kaplan, there is a risk that the reason for ineligibility may be associated with outcome in such a way as to cause bias (e.g. a patient has a prognosis significantly different from that desired by the trial protocol).

Wherever a PPA analysis is used, an equivalent ITT analysis will also be presented. The choice of whether PPA or ITT will form the primary analysis sample will be pre-specified, depending upon the extent of non-compliance; see next section.

5.8. Analysis models

Primary model

A Cox model will be used to generate an estimate of the HR between randomised groups, to be compared to the relevant critical value specified *a priori* in the MAMS design (see Sections 5.5.1 and 5.5.2). 95% confidence intervals will be presented for the HR. To improve precision, the model will be adjusted for the following factors used in the minimisation process at randomisation (see also Section 3):

- Site of primary tumour (Right colon; Left colon; Rectum)
- WHO performance status (0; 1; 2)
- Disease assessment from baseline CT scan using RECIST v1.1 (must be after 16 weeks of first-line therapy and within 4 weeks prior to randomisation) (Complete response; Partial response; Stable disease)
- Number of metastatic sites (0; 1; 2+)
- Type of first-line therapy:
    - Fluoropyrimidine drug (5FU; Capecitabine; Neither)
    - Use of oxaliplatin/irinotecan (Both; Oxaliplatin only; Irinotecan only; Neither)
    - Use of monoclonal antibody (Cetuximab/Panitumumab; Bevacizumab; None)

Given the number of sites participating in FOCUS4 (>100), randomisation site (also a minimisation factor) will not be used in this model.

For FOCUS4-N, randomisation was also minimised on biomarker cohort. The precise categories changed over the lifetime of the FOCUS4 trial in response to adaptations to the protocol and to the constituent trial cohorts. Therefore, the following "hybrid" categories will be used for adjustment:

- BRAF mutation
- PIK3CA mutation or PTEN loss (if BRAF wild-type)
- KRAS or NRAS mutation (if BRAF and PIK3CA wild-type, and PTEN other than loss)
- All wild type (that is, BRAF, KRAS, NRAS, PIK3CA wild-type, and PTEN other than loss)
- Non-stratified, if not allocated to one of the above categories (e.g. due to test failures).

Furthermore, the primary analysis model for FOCUS4-N will also be stratified (rather than adjusted) for the following two timepoints:

- Timing of randomisation with respect to closure of Trial D on 18th March 2016
- Timing of randomisation with respect to removal of exclusion criterion for high platelets as of 30th June 2017.
    - Note that Trial C was opened as of this same date, so that for Trial N this factor accounts simultaneously for the change to the platelet count criterion and to the opening of Trial C.

Alternative primary model (if necessary)

For some early interim analyses when few patients and events are available, the Primary Model described above may not converge successfully. In this instance, the inverse-probability weighting method will be used as for adjustment of minimisation factors in this Primary Model. (Note that this scenario may coincide with use of a per-protocol analysis sample as described in Section 5.6.)

A logistic regression model is first fitted to the analysis sample to calculate the propensity of being randomised into each arm of the Trial (i.e. predicted probability) on the basis of the minimisation factors listed above. (If certain factors, e.g. with low cell counts, need to be omitted in order for the model to fit, this should be made clear; the number of such factors should however be minimal.) The predicted probabilities from this model are then subtracted from 1 for patients on the control arm, before the reciprocal of all values is taken to form the inverse-probability weights. An unadjusted Cox model estimating the treatment HR is then run, incorporating these weights into the survival structure of the data (e.g. using the "stset" command in Stata). Note that although the point estimate may immediately be compared with the critical value from the MAMS design as before, the resulting variance is known to be overestimated (in theory, the outputted variance should be similar to that of the unadjusted model, whereas it should be smaller). Hence, for presentation purposes a suitable 95% confidence interval should be obtained using the bootstrap method with bias-correction and acceleration factor, with 1,000 replications.

Secondary models

Regardless of whether the default or alternative primary analysis model is used, a completely unadjusted Cox model estimating the treatment HR and 95% confidence interval, together with a log-rank test, will be presented to check that the point estimate is broadly consistent between the models.

Additionally, if there are adequate numbers of events to justify it, a secondary adjusted model may be fitted. It is not recommended that the secondary adjusted estimate of treatment HR is used to make the decision of whether or not to continue, but it may lead to recommendations for what data to present at subsequent IDMC meetings. The 95% confidence interval from the model will be presented along with the treatment HR. Again, in case of non-convergence, the inverse-probability weighting propensity score method may be used, in which case a suitable 95% confidence interval for the treatment HR should be obtained using the bootstrap method (see above).

Adjusted factors for this model are the minimisation factors listed above, plus the following additional prognostic factors:

- Resection status (Resected primary; Unresected primary; Unresectable local recurrence)
- Timing of metastases (Synchronous; Metachronous)
- Alkaline phosphatase (continuous)
- White blood cell count (continuous)
- Age of biomarker panel at randomisation (continuous, log-transformed)
- Aspirin use (yes; no)

Extra exploratory sensitivity analysis (discretionary only)

If any unexpected imbalances between randomised groups are noted during data analysis, the trial statistician may use their discretion to perform an extra sensitivity analysis including the unexpectedly imbalanced variable(s) in addition to all the above factors (using the propensity score method if necessary). This will be made clear within the presentation of data to the IDMC.

5.9. Exploratory Analyses of Efficacy

Analyses described in this section will use only the Primary Outcomes defined in Section 5.4.

Pre-specified subgroup analyses

All subgroup and interaction analyses, as described below, are considered exploratory only. In general, treatment effect estimates will be presented within each subgroup category. A test for interaction between treatment and subgroup category will be performed using a Cox regression model with main effects for treatment and subgroup category, and with treatment-interaction terms for each subgroup category. Note, however, that for smaller cohorts the amount of data within

certain subgroups may be insufficient to obtain a reliable model estimate of treatment effect and/or interaction.

For all trial cohorts, the following minimisation factors used at randomisation will be analysed as exploratory subgroup analyses:

- Site of primary tumour (Right colon; Left colon; Rectum)
- WHO Performance Status (0/1/2)
- 16-week CT scan result (Stable disease; Partial response; Complete response)
- Number of metastatic sites (0/1/2+)
- First-line chemotherapy regimen
    - Fluoropyrimidine (5FU; Capecitabine; Neither)
    - Oxaliplatin/irinotecan (Both; Ox only; Ir only; Neither)
    - Monoclonal antibody (Cetuximab/Panitumumab; Bevacizumab; None)

For FOCUS4-C, the following additional subgroup analyses will be undertaken:

- PIK3CA mutation vs wild-type
- CMS sub-type (if available)

For FOCUS4-N, the following additional subgroup analyses will be undertaken:

- BRAF mutation status (mutation vs wild-type)
- PIK3CA mutation and PTEN loss status (mutation or loss vs neither mutation nor loss)
- RAS mutation status (mutation vs wild-type)
- RAS+p53 “double mutation” status (mutation vs wild-type; c.f. FOCUS4-C)
- “All wild type” status (BRAF, RAS, PIK3CA wild-type and no PTEN loss vs any mutation or PTEN loss)
- “Non-stratified” status (non-stratified vs other)
- Platelet count (high vs low at 400 threshold, only in the subgroup of patients randomised after the inclusion criterion was relaxed)

Analysis of Response

According to RECIST criteria, pre-randomisation CT scans will be classified as either Progressive Disease (PD; in which case they would not be randomised); Stable Disease (SD); Partial Response (PR) or Complete Response (CR). Since disease response is one of the minimisation criteria, we expect the distribution of response to be balanced by treatment arm.

For those patients with measurable disease at their pre-randomisation CT scan, the following analyses will be performed by treatment arm:

- 16-week response
- 16-week waterfall plot, based on % change in primary tumour diameter from pre-randomisation CT scan
- Best response at any time
- Duration of response

Swimmer plots will also be compiled, separately by treatment arm, showing disease progression of each individual patient from randomisation (or, more precisely, from pre-randomisation CT scan) to either leaving the trial, disease progression, or death.

FOCUS4-C only: Comparison of doses for AZD1775

At the start of FOCUS4-C, the standard dose of AZD1775 was at 250mg, with a higher dose of 300mg only permitted at Level 3 sites. During the course of the trial, a series of tolerability reviews were undertaken to see whether the standard dose of 250mg could be safely increased to 300mg. The IDMC ultimately recommended that the dose increase should proceed, and this was ratified by the TSC and implemented by the trial team. This has resulted in the AZD1775 arm having roughly half of patients commencing treatment at 250mg, and roughly half at 300mg. The primary analysis will pool both doses indiscriminately, but a secondary analysis will also be presented for each dose separately. Since at least some patients could commence trial treatment on either dose throughout the duration of the trial, it is not possible to compare each dose with contemporaneous control arm data. Instead, therefore, the Primary Model described in Section 5.8 will be fitted with a three-level treatment variable (coded 1 = Active Monitoring, 2 = AZD1775 at 250mg, 3 = AZD1775 at 300mg) and the hazard ratios for each dose compared to Active Monitoring will be presented, together with Kaplan-Meier curves. There will not have adequate power to make a formal comparison of dose efficacy, so this will be an exploratory analysis that sits alongside the presentation of safety and toxicity data which will also be broken down by dose.

Indirect analysis of data from FOCUS4-C and FOCUS4-N

Recall that in order to be eligible for randomisation into the FOCUS4-C sub-trial, patients must have tumours with both RAS mutation and p53 mutation. There will also be a subset of patients randomised into the FOCUS4-N sub-trial with these mutational characteristics, who did not enter FOCUS4-C due to patient or clinician choice, or because FOCUS4-C was not open for randomisation at their site or at the appropriate time. Therefore, since both sub-trials used the same control arm (Active Monitoring) there is an opportunity to indirectly compare research treatments, and thereby explore the efficacy of AZD1775 with capecitabine for patients with RAS+p53 mutated tumours.

In order to perform this analysis, simplified datasets will be created for each of the two sub-trials (that is, all FOCUS4-C patients; and the subset of FOCUS4-N patients with a RAS+p53 double mutation): containing treatment allocation, PFS survival time and event status, plus minimisation factors and adjustment covariates as described in Section 5.8. All included data will be subject to the same numerical coding and/or measurement unit in both datasets. These datasets will then be appended to form a single, larger dataset with one observation (“row”) for each patient randomised either into the FOCUS4-C, or into FOCUS4-N with a RAS+p53 double mutation.

The analysis will then proceed as described as for "**Pre-specified subgroup analyses**" above, with the outcome of PFS; that is, using a Cox regression model with main effects for treatment (active drug vs Active Monitoring) and trial cohort (FOCUS4-C vs FOCUS4-N), and a treatment-interaction term. In addition, a Kaplan-Meier plot will be produced to compare the four treatment arms (AZD1775, Capecitabine, plus the two Active Monitoring arms).

5.10. Secondary outcomes

Secondary efficacy outcomes

Time to second progression

Evaluation of disease control from CT scans at 8-weekly intervals

Trial treatment and compliance

Treatment and compliance in both arms will be described at each interim analysis using the following summary data:

- Median (IQR) number of cycles received, by arm
- Median (IQR) dose intensity, defined as the ratio of observed to expected dosage (for the cycles actually received) multiplied by the ratio of observed to expected number of cycles received (for the time on trial)
- Pill taking compliance summarised by the overall number of missed doses presented as the median (IQR) number of pills returned.

Toxicities and Symptoms

Toxicities and symptoms are assessed every 4 weeks (or 8 weeks for FOCUS4-N) throughout the FOCUS4 treatment period using NCI common terminology criteria (NCI CTC version 3.0). The "worst" CTC grade experienced during the previous 4 weeks is reported. The primary analysis for toxicities will be on a Per-Protocol basis, but ITT analyses will also be presented. For FOCUS4-C, toxicities will also be presented separately by AZD1775 dose (250mg vs 300mg; see Section 5.9).

At each interim analysis, toxicity and symptom data will be converted into a binary variable representing whether each patient has experienced a CTC Grade 3+ toxicity/symptom at any time since randomisation. Numerators, denominators and percentages will be presented, by treatment allocation, for each individual toxicity/symptom and for toxicities/symptoms grouped by body system as follows:

01 = Nausea	*09 = Anaemia*	*18 = Dry eyes*
02 = Vomiting	*10 = Neutropenia*	*19 = Photophobia*
03 = Diarrhoea	*11 = Thrombocytopaenia*	*20 = Blurred vision*
04 = Stomatitis	*12 = Hyperbilirubinaemia*	*21 = Conjunctivitis*

05 = Dry skin	*13 = Transaminitis*	*22 = Corneal ulceration*
06 = Skin Rash	*14 = Hypomagnesaemia*	*23 = Fatigue*
07 = Acne	*15 = Cardiac toxicity*	
08 = PPE	*16 = Pneumonitis*	
	17 = Infection	*26 = Other*

Serious adverse events

Serious Adverse Events (SAEs) are reported to the MRC CTU by investigators within one working day of their becoming aware of the event. SAEs will be described at each interim analysis as follows:

- Raw frequencies and percentages (including repeats within patients), by treatment arm
- Raw listing of all relevant data (including treatment arm) if required.

Quality of Life (QoL)

Quality of Life (QoL) data will only be measured by EuroQol-5D, version 2.

QoL will only be assessed in any molecular trial that continues into Stages III or IV. However, QoL data may be collected at earlier stages if it is deemed to be important for that specific trial.

QoL will be measured from randomisation in all FOCUS4-N patients. ANCOVA methods will be used to analyse any difference in QoL scores between randomised groups adjusted for baseline scores and minimisation factors (see Section 3).

Full details of the analysis of QoL data will be given at a later date, either in a subsequent version of this document, or in a separate Analysis Plan document.

6. Data and analysis completion schedule

Depending on the cleanliness of the dataset, preparation and cleaning of the dataset is expected to take ~2 weeks. Analysis of the main results including the primary outcome will take a further 2 weeks and analysis of the secondary outcomes a further 2 weeks. Thus, the process for completing an interim analysis is likely to take ~6weeks. The trial statistician will need to anticipate when the required number of events is likely to occur for each interim analysis. This will trigger commencement of the data cleaning for that interim analysis.

7. Signatures of Approval

Date: 2nd November 2020

Version: 3.0

Signatures:

Name:	Trial Role:	Signature:	Date signed:
Max Parmar	MRC CTU Director		3rd Nov 2020
Rick Kaplan	MRC CTU Co-investigator		3rd Nov 2020
Louise Brown	Senior statistician		3rd Nov 2020
David Fisher	Trial statistician		3rd Nov 2020

Appendix: Analysis plan for testing agents in biomarker-negative cohorts

Proposed trials in biomarker-negative cohorts:

- Following IDMC approval for an agent to pass Stage II, a new trial will be designed to test the agent in future patients not selected for that biomarker cohort (biomarker-negative). The new trial will probably use a similar randomised MAMS approach to the original biomarker-positive cohort, with similar rules for early stopping for lack of sufficient activity, albeit seeking a possibly smaller treatment effect (ie. a bigger trial).
- Depending on the contemporaneous status of the other cohorts, the biomarker-negative cohort may encompass patients from a different cohort in which the trial for a selected agent is not currently open. In other circumstances, the biomarker-negative cohort may be assembled by diverting patients from the other cohorts that have different biomarker profiles. However, if there is good published evidence that the agent is unlikely to have activity in a specific molecular subtype, eg. KRAS mutant, then this group may be excluded from the biomarker-negative trial.
- The precise trial design for biomarker-negative cohorts, including details of sample size, timings and analyses, will be agreed (in advance of relevant analysis) after discussion with the IDMC, TMG, TSC and relevant industrial collaborator(s).

Analytical methods for biomarker-negative cohorts:

A separate SAP will be developed for each biomarker-negative cohort, but there are likely to be common methods used for biomarker-positive and -negative cohorts. For example, both will use a Cox survival model for testing the randomised treatment effect, probably adjusting for the same covariates (see Section 5.7).

Since treatment efficacy is tested in both biomarker-positive and -negative cohorts under a formal pre-specified trial design, the primary interpretation of the two results should be as two separate but related trials, with the effect size and statistical significance of each made clear. A forest plot of treatment effects may be presented to aid interpretation of whether a significant effect is seen for each group. Three scenarios may emerge:

1. Both biomarker-positive and -negative patients derive significant benefit from the drug. This will indicate that the biomarker is not helpful in discriminating patients into those who do or do not respond to treatment.
2. The biomarker-positive patients do derive a statistically significant benefit but the biomarker-negative patients do not derive a statistically significant benefit (a lower HR threshold will have been agreed in advance and used for the sample size calculations in the biomarker-negative trial).
3. The biomarker-positive patients do derive a statistically significant benefit and the biomarker-negative patients derive some benefit in terms of a reduced HR but the 95% CI does not exclude the possibility of no benefit.

Alternative methods may be developed for presenting the findings for biomarker-positive and -negative patients, e.g. Venn diagram presentation of treatment effects in biomarker-positive and -negative patients by randomised group.

However, as all patients will have been investigated using the same overall Master and Trial Protocols, it may be appropriate (e.g. given scenario 3 above) to perform an exploratory meta-analysis of treatment effects stratified by biomarker status. In this case:

- The rationale for doing so will be given in the SAP for the relevant biomarker-negative cohort, together with a description of the circumstances under which it would be done, e.g. that the point estimates of effect in the two cohorts are sufficiently similar
- The meta-analysis would probably use a fixed-effects (due to the fact that the number of treatment effects is small, and fixed in advance) inverse-variance model, with heterogeneity assessed using the Q statistic. The I-squared statistic (Higgins) may also be generated with confidence limits estimated using the "generalised Q" method (Viechtbauer; Bowden).

Careful interpretation of results across biomarker cohorts will be required as HRs will be based upon averages and some patients may still derive benefit from a particular agent even if their biomarker cohort does not.

Acknowledgement

This Statistical Analysis Template is constructed with the aid of an original document by

Matthew Sydes.