
# STATISTICAL ANALYSIS PLAN

**Version 1.0 4-Dec-2017**

| Author | Position | Signature | Date |
|---|---|---|---|
| Wolfgang Stöhr | Trial Statistician | | 4 Dec 2017 |
| **Approved by** | **Position** | **Signature** | **Date** |
| Abdel Babiker | Project Lead & Statistician | | 6 Dec 2017 |
| Sarah Fidler | Chief Investigator | | 05.12.2017 |

**Revision History**

| Version | Author | Date | Reason for Revision |
|---|---|---|---|
| 0.1 | A Babiker, W Stöhr | 24-Oct-2014 | As in protocol version 1.0 |
| 0.2 | W Stöhr | 09-Mar-2017 | Initial full draft |
| 0.3 | W Stöhr | 16-Mar-2017 | Reviewed by A Babiker |
| 0.4 | W Stöhr | 20-Jul-2017 | Reviewed by S Fidler, S Pett, L Dorrell, A Fun, J Frater |
| 1.0 | W Stöhr | 04-Dec-2017 | Final version, after further review by A. Babiker, S Fidler, S Pett, L Dorrell, A Owen, J Frater.<br>Major changes:<br>- Primary Endpoint: Clarifying handling of undetectable and missing values<br>- Viral Outgrowth: Clarifying analysis method<br>- HIV-specific T cell responses: Changed method from Elispot to ICS |

# TABLE OF CONTENTS

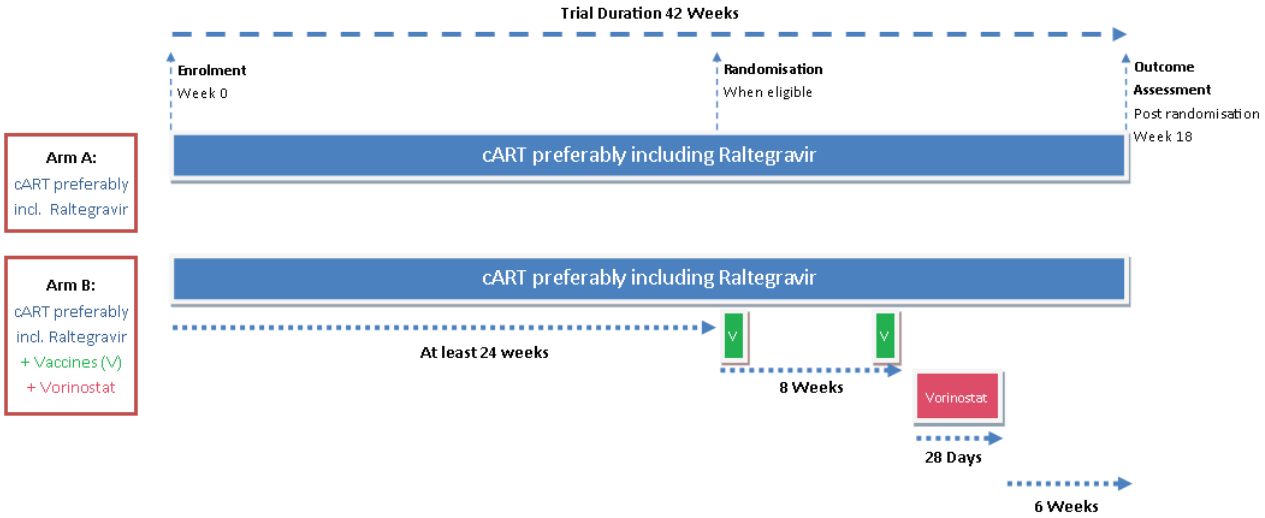
# 1 OVERVIEW OF RIVER

## 1.1 SUMMARY

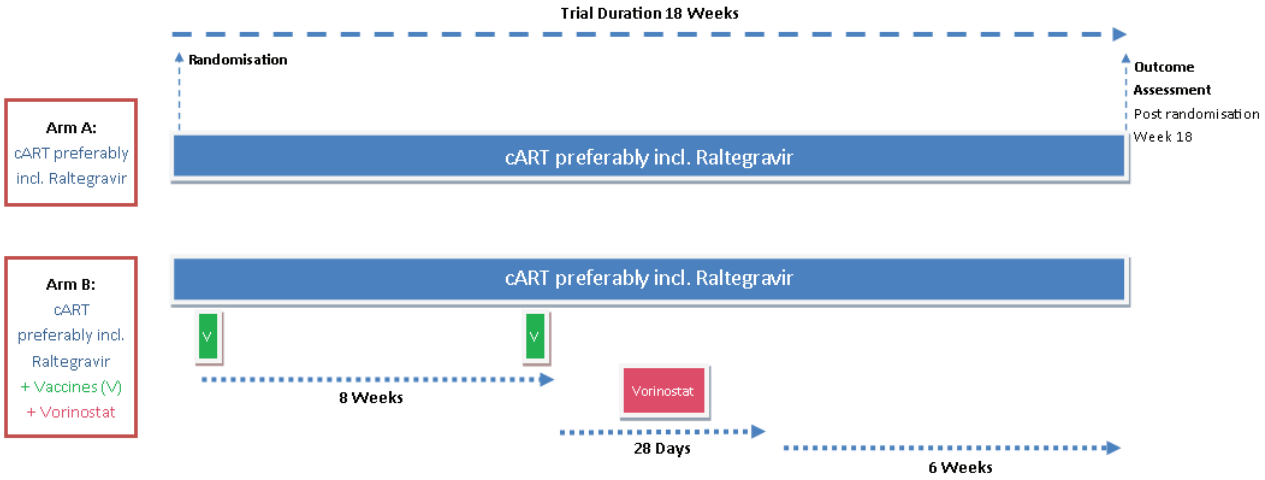
| TRIAL NAME | RIVER (RESEARCH IN VIRAL ERADICATION OF HIV RESERVOIRS) |
|---|---|
| Interventions to be compared | This study is a two-arm prospective randomised controlled trial comparing:<br>**Arm A:** 4-drug cART including raltegravir (control)<br>**Arm B:** 4-drug cART including raltegravir plus ChAdV63.HIVconsv prime and MVA.HIVconsv boost vaccines; followed by a 28-day course of vorinostat (10 doses in total). |
| Study Hypothesis | In primary HIV infection, a combination of immediate cART, immunisation and latency reactivation using the HDACi vorinostat will confer a significant reduction in the reservoir when compared with cART alone. |
| Participants | The study aims to enrol 60 individuals across 6 UK collaborating clinical centres, from two different strata:<br>**Stratum I:** Recently diagnosed with primary HIV-1 infection [see protocol for definition]. Participants must be enrolled within 4 weeks of a confirmed diagnosis of primary HIV-1 infection. Eligible participants are enrolled at week 0 when combination ART (cART) begins. Randomisation of participants occurs after assessment of eligibility at week 22.<br>**Stratum II:** Previously diagnosed with primary HIV-1 infection. Confirmed diagnosis of primary HIV-1 infection must be within a maximum of 2 years prior to randomisation. Participants must have started at least 3-drug ART within 4 weeks of a confirmed diagnosis of primary HIV-1 infection and remained on ART since starting. Randomisation of participants occurs within 2 weeks of the assessment of eligibility at screening. |
| Randomisation | The participants enrolled in the trial are randomised in a 1:1 ratio to one of two intervention arms, with randomisation stratified by the two strata defined above. |

## 1.2 TRIAL SCHEMA

### Stratum I - Recently diagnosed




### Stratum II - Previously diagnosed



## 1.3 OUTCOME MEASURES

### 1.3.1 PRIMARY OUTCOME MEASURE

The primary outcome measure is total HIV DNA from CD4 T-cells averaged across post-randomisation weeks 16 and 18.

### 1.3.2 SECONDARY OUTCOME MEASURES

- Clinical and laboratory adverse events, all grades, including SAE
- Further assessment of the HIV reservoir e.g. HIV integrated DNA; HIV cell associated RNA; plasma HIV RNA measured with an ultra-low copy assay i.e. with a threshold of <1 copy/ml, viral outgrowth assays
- Studies of immune function including measuring the latently-infected resting memory T-cells and HIV-specific T cell responses (in protocol vs 5.0: cytotoxic immune responses)
- Changes in inflammatory biomarkers

## 1.4 SAMPLE SIZE

The sample size calculation is based on the primary endpoint which will be analysed on a $\log_{10}$-scale. The following assumptions are made:

- The combination intervention will confer a 50% reduction in total HIV DNA when compared with cART alone. On a $\log_{10}$-scale this corresponds to a difference between the two arms of $\log_{10}(2)$ (approximately 0.3). The effect is assumed to be the same in both strata (recently and previously diagnosed patients).
- Standard deviation (SD) is 0.4 for a single measurement in both arms based on two publications:
  - Reported on data from 31 participants treated with HAART at PHI in the French PRIMO study. The median (IQR) at 6 and 12 months after starting treatment were 2.30 (2.10, 2.70) and 2.10 (1.80, 2.40) $\log_{10}$ DNA copies/$10^6$ PBMC respectively suggesting a SD of 0.45 at both time points[57].
  - Reported on 8 participants with PHI. At week 52, the median (IQR) was 494 (250 - 694) for total HIV-1 DNA copies/$10^6$ CD4+ T-cells suggesting a SD (in log scale) of 0.33[49].
- 1:1 allocation of participants in the two study arms.
- Method of analysis: comparing the treatment arms in terms of absolute HIV total DNA level at post-randomisation weeks 16 and 18 adjusted for baseline (here: randomisation) level (analysis of covariance; ANCOVA); one baseline measurement and two follow-up measurements (at PR16 and 18) will be taken for an individual participant.
- A correlation coefficient of 0.5 between a baseline measurement and a PR 16/18 measurement, and a correlation coefficient of 0.7 for measurements at PR 16 and 18.
- Two-sided $\alpha = 0.05$ for the null hypothesis that there is no difference between the two arms in the primary endpoint.

Under the above assumptions, a sample size of 52 individuals would provide 94% power to detect a 50% reduction in HIV total DNA (86% power for a 45% reduction).

Assuming that some individuals enrolled will not reach visits PR 16/18 (due to failing the eligibility criteria for randomisation in stratum I, or withdrawal) it is planned to enrol a total of 60 individuals overall (both strata combined).

# 2 DEFINITIONS

## 2.1 DEFINITION OF BASELINE

Baseline is defined in this trial as the time of randomisation. Baseline values used to define changes in immunological, virological and other measurements over time for each participant are defined as the measurements at randomisation. If a measurement at randomisation is not available, then the measurement at week 22 (stratum I) or screening (stratum II) closest to randomisation (but within 14 days before) will be used. If no measurement is available at or prior to randomisation then any measurement taken up to 7 days post-randomisation will be used providing this is before the start of vaccination.

## 2.2 DEFINITION OF UNDETECTABLE HIV-RNA

Undetectable plasma HIV RNA is defined <200 copies/mL for the Taqman Roche 2.0 assay, and as <50 copies/mL for all other assays.

## 2.3 ADVERSE EVENTS

Adverse events will be reported on the case report form. Prior to randomisation only grade 3 and 4 adverse events and Serious Adverse Events (SAE) are collected. Post-randomisation, all adverse events will be reported including grade 1 and 2 events, and other notable events (see below).

All clinical events reported during the trial will be reviewed by the Trial Physician (unblinded). Adverse events will be classified by system organ class and lower level term according to the version of MedDRA® implemented at the start of the trial (version 19; the same version will be used throughout).

Adverse events (clinical and laboratory) will be graded using the 2004 (amended 2009) Division of AIDS toxicity grading table (see Appendix III of RIVER protocol).

Serious adverse events will be defined according to principles of GCP.

Other notable events, to be reported on designated CRFs: Vaccine-related events (see Appendix IV of the RIVER protocol), Cancer, Pregnancy in a partner.

Solicited adverse events following the receipt of each vaccine will be recorded on a CRF during the first 60 mins by clinical staff, and then daily by the participant on a diary card for 3 days.

## 2.4 COUNTING OF EVENTS

SAEs will be analysed as episodes, with all components of the same clinical SAE presented as one episode. Analyses of grade 3 or 4 AEs will consider each graded component as separate events. All SAEs are also reported as adverse events; all grade 4 adverse events are reported as SAEs (life-threatening by definition).

## 2.5 OTHER DEFINITIONS

Glomerular filtration rate will be estimated by Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) creatinine based GFR estimating equation [1].

# 3 ANALYSIS PRINCIPLES

- The trial will be primarily analysed as intention-to-treat including all randomised participants. Based on the consideration that the trial is evaluating the efficacy, as measured by the impact on total CD4+ HV DNA of a novel intervention, sensitivity analyses will also be performed for immunological/virological outcomes excluding participants with incomplete intervention (temporary/complete stop of ART, vaccination, or vorinostat).
- Randomisation arms will be compared using data between baseline and the post-randomisation week 18 visit (inclusive). Data collected in the period between enrolment and randomisation will be reported for stratum I as baseline data.
- For continuous variables, the following will be presented by scheduled visits and by randomised group:
  - number of participants with a measurement
  - mean (SD) or median (IQR) of absolute values and of changes in absolute values from baseline
  - mean changes (plus point-wise 95% CI) from baseline in each randomised group, calculated using normal regression of absolute values, adjusting for the baseline measurement.
  - normal generalized linear regression models (using GEE) will be used for a global test of difference between treatment groups across all visits

  If positively skewed, data will be log-transformed before the analysis. If parametric models/tests do not seem to be appropriate, rank tests will be performed.
- Binary and categorical variables will be tabulated, overall and by randomised group. Differences between groups at particular time-points will be tested using chi-squared tests (or exact tests if appropriate), and logistic regression models for adjusted analyses. GEE models will be used for a global test of difference between treatment groups across all visits.
- Time-to-event outcomes will consider time from baseline to the event date, using Kaplan-Meier estimation. For participants who have not reached the event in question, the date of last clinic visit will be used as a censoring time. Differences between groups in time-to-event outcomes will be tested using a log-rank test and Cox proportional hazard regression models (stratified by the stratification factors).
- Primary analyses will adjusted for membership of stratum I versus stratum II; secondary analysis will be conducted without adjustment.
- Multiple imputations will be performed to account for missing data in the primary outcome measure (see below). Analyses of secondary endpoint, unless otherwise stated, will be based on available data only. Reasons for missing efficacy data will be described.
- All estimates including differences between randomised groups will be presented with a 2-sided 95% confidence interval. All statistical tests will be 2-sided. P values will be given to 2 decimal places if ≥0.10, otherwise to 1 significant figure.

# 4 ANALYSIS DETAILS

## 4.1 ENROLMENT AND ELIGIBILITY

The following results will be presented overall, and by stratum:

- Total screened for enrolment and enrolled by site, with dates of first and latest enrolment
- Reasons for not enrolling screened patients
- Enrolment over calendar time: cumulative enrolment; enrolment by calendar month
- Eligibility: number (%) and reasons for any ineligibilities (i.e. enrolled although eligibility criteria violated)

## 4.2 RANDOMISATION AND ELIGIBILITY

The following results will be presented overall, by stratum, and by randomised group:

- Total screened for randomisation by site
- Total randomised by site, with dates of first and latest randomisation
- Reasons for not randomising screened patients
- Randomisation over calendar time: cumulative randomisation; randomisation by calendar month
- Eligibility: number (%) and reasons for any ineligibilities

## 4.3 PATIENT CHARACTERISTICS

The following baseline characteristics will be presented overall and by randomised group. For the grouping of quantitative variables, categories will be chosen after univariate inspection of the data, ensuring that a reasonable number of participants are represented in each category.
Before completion of randomisation, the same characteristics will also be reported for all patients enrolled but not by randomised group.

- Stratum: number (%) stratum I, stratum II
- Age: median (IQR), range; distribution in categories
- Sex: number (%) male, female
- Ethnicity: number (%) White, South Asian, South East Asian, Hispanic/Latino, Black African, Black Caribbean/American, mixed ethnic group, other
- Mode of HIV infection: number (%) sex between men, sex between men and women, injection drug use, blood products, unknown, other
- CD4 count (current): median (IQR), range; distribution into categories
- HIV-RNA: number (%) undetectable; distribution into categories
- Criteria for PHI diagnosis: number (%) positive HIV-1 serology, positive p24, HIV RNA or proviral DNA, 4$^{th}$ generation HIV test, western blot, RITA only.
- eGFR: median (IQR), range
- QTc interval: median (IQR), range
- Haematology/biochemistry/electrolytes/glucose/lipids/urine analysis: median (IQR), range of test results

## 4.4 DESCRIPTION OF FOLLOW-UP

- Time between randomisation and last follow-up (weeks): median (IQR), range
- Attendance of scheduled visits, by visit: number (%) missed visits. Denominator to include any patients withdrawn or lost to follow-up, but not any known to have died.
- Additional visits compared with schedule: number (%). Denominator to include visits for any patients withdrawn or lost to follow-up, but not any known to have died.
- Early end of trial participation: number (%); description of reasons.
- End of treatment without end of trial participation: number (%).

For reports while the trial is ongoing:

- Time between enrolment and randomisation (weeks) in stratum I: median (IQR), range
- Lost to follow-up: number (%) formally defined as lost to follow-up (not seen in the clinic for more than 3 months), subdivided by formal ending of trial participation (yes/no)
- Transfer to a new site: number (%)
- Last attended scheduled visist: number (%)

## 4.5 DESCRIPTION OF TRIAL TREATMENT

**ART pre-randomisation:**

- First regimen: number (%) per regimen
- Time from PHI diagnosis to starting ART: median (IQR), range; distribution into categories ≤1 week, 1-2 weeks, 3-4 weeks.
- ART pre-PHI diagnosis: number (%) pre-exposure prophylaxis, post-exposure prophylaxis
- Time from starting ART to randomisation: median (IQR), range; distribution into categories ≤24 weeks, 24-48 weeks, 48-96 weeks, >96 weeks.

**ART from randomisation:**

- ART regimen at randomisation: number (%) 3-drug/4-drug regimen, number (%) on integrase inhibitor, number (%) per regimen
- ART changes: number (%) making changes, reasons for changes, drugs changed to; time to first change
- ART discontinuation: number (%) with permanent discontinuation of all ART (>30 days), description of reasons
- Regimen at last follow-up: number (%)
- Drug class exposure: number (%) ever exposed to 2, 3, 4 classes of drugs (i.e. cumulative to date)
- Non-adherence: any dose of ART missed since last visit: number (%); any dose missed in last 7 days: number (%), overall and by drug; how many doses of ART missed in last 7 days: number.

**Vaccines (participants randomised to arm B only):**

- Not received ChAd vaccination: number (%); description of reasons
- Not received MVA vaccination: number (%), description of reasons

**Vorinostat (participants randomised to arm B only):**

- Not started vorinostat: number (%); description of reasons
- Treatment deviations/modifications: number (%) with dose modification, missed doses, treatment discontinuation; description of reasons

## 4.6 PRIMARY ENDPOINT

The primary outcome measure is based on total HIV DNA from CD4 T-cells (copies/million CD4+T cells) which is measured at enrolment (in recently infected patients), baseline, PR week 08-3, PR week 12-2, PR week 16, and PR week 18.

The primary endpoint is total HIV DNA averaged across post-randomisation weeks 16 and 18. It will be analysed on a $\log_{10}$-scale. Treatment arms will be compared in terms of absolute total HIV DNA levels at post-randomisation weeks 16 and 18 adjusted for the baseline (i.e. randomisation) level and by stratum using analysis of covariance.

Each sample of a particular time-point is measured in triplicates. Any result below the lower detection limit (LDL) - including undetectable DNA but NOT missing values - will be replaced by LDL/2. A value of 5 HIV copies/well will be adopted as lower limit of detection.

**Handling of missing total HIV DNA measurements**

If either the PR week 16 result or the PR week 18 result is entirely missing but not both, the primary endpoint consists of the single available result. If total HIV DNA is missing at both PR-16 or PR-18, or at baseline, an imputation method will be used to estimate missing values. Predictors may include stratum, total HIV DNA from previous time-points and other factors associated with total HIV DNA eg CD4.

**Sensitivity analyses for the primary endpoint**

The following sensitivity analyses for the primary endpoint will be performed:

- With the per protocol population, i.e. excluding participants with incomplete intervention (temporary/complete stop of ART, vaccination, or vorinostat).
- Complete case analysis: excluding participants with missing primary endpoint

## 4.7 SECONDARY ENDPOINTS: EFFICACY

- **HIV integrated DNA**:

  Measurements (copies/million CD4+T cells) at enrolment (in recently infected patients), baseline, PR week 08-3, PR week 12-2, PR week 16, and PR week 18.

  Analysed as for continuous outcomes as specified above. Treatment arms will be compared in terms of absolute integrated HIV DNA levels at post-randomisation weeks 16 and 18 adjusted for the baseline level and by stratum.

- **Plasma HIV RNA measured with an ultra-low copy assay:**

  Number (%) with undetectable HIV-RNA at enrolment (in recently infected patients), baseline, PR week 08-3, PR week 09-2, PR week 12-2, PR week 16, and PR week 18; analysed as for binary outcomes as specified above.

- **Replication competence (Viral Outgrowth assay):**

  Measurements (infectious units per million cells; IUPM) at baseline and PR week 16. Analysed as for binary and continuous outcomes as specified above. The results are expected to be "negative" (i.e. below the lower limit of detection) in a sizable proportion of samples. Therefore, the treatment arms will be compared as follows:

  1) Primary analysis:
     a) Comparison of the proportion of patients with a "negative" result using logistic regression. This analysis includes all patients with a valid PR-16 result.
     b) Comparison of IUPM in patients with "positive" results using median regression. Patients with a "negative" PR-16 result will be excluded.
  2) Secondary analysis:
     Comparison of IUPM using a linear model appropriate for left-censored data. This analysis includes all patients with a valid PR-16 result, "positive" or "negative".

  In all analyses, patients with missing PR-16 samples or invalid PR-16 results (due to assay failure) will be ignored. Analyses will be adjusted for stratum and baseline viral outgrowth. Missing baseline values will be imputed; predictors may include stratum, and other factors associated eg CD4, HIV-DNA.

- **HIV-specific T cell responses (both CD8+ and CD4+ T cell responses):**

  Measurements using intracellular cytokine staining (ICS) at enrolment (in recently infected patients), baseline, PR week 09-2, PR week 12-2, and PR week 16. For both CD8+ and CD4+ T cells, the following markers will be measured: CD107a+, CD154+, IFNg+, IL2+, and TNFa+.

  For each marker, the magnitude of a response is defined as the number of positive cells as percentage of their parental populations, after subtracting the negative control. For technical reasons, HIV antigens will be divided in non-overlapping pools values for each sample; responses will be added up for each marker.

  At each time-point, the magnitude of responses, and the change from baseline will be analysed as for continuous outcomes as specified above.

- **CD8+ T cell antiviral activity (viral inhibition)**

  Measurements using a flow cytometer-based assay [2] at enrolment (in recently infected patients), baseline, PR week 09-2, PR week 12-2, and PR week 16.

  CD8+ T cell antiviral suppressive activity will be expressed as percentage inhibition and determined as follows: [(fraction of p24 + cells in CD4 + T cells cultured alone)– (fraction of p24 + in CD4+ T cells cultured with CD8+ cells)]/(fraction of p24 + cells in CD4 +T cells cultured alone) × 100.

  CD8+ T cell antiviral suppressive activity will be analysed as described above for continuous outcomes.

- **Inflammatory biomarkers:**

  For example CD4:8 ratio, immune exhaustion/activation markers, IL-6 and d-dimer, neuro-inflammation markers. Analysed as for continuous outcomes as specified above.

- **HIV cell associated RNA (arm B only):**

  Measurements (copies/million 18S copies) pre and 2 hours post vorinostat at visits PR week 08-3, PR week 09-2, and PR week 12-2. Comparison of pre and post vorinostat values will performed using paired t-tests (or paired rank-tests is appropriate).

## 4.8 SECONDARY ENDPOINTS: SAFETY

Adverse events pre-randomisation wil be presented overall. Adverse events post-randomisation wil be presented by randomised arm. For all adverse event types, the total number of events, the number of participants with at least one event, and the maximum grade per participant will be given. Analysis of adverse event outcomes will be based on the number ever experienced a type of adverse event which will be compared as for binary outcomes. Timing of events, and recurrent events will also be described.

Where relevant and established, events will also be presented by severity grade, and by relationship to study treatment (definitely, probable, possible, unlikely, unrelated).

The following adverse event outcomes will be analysed:

- Grade 3 or 4 adverse events
- Notable events
- Treatment-modifying adverse events (any grade)
- Post-vaccination solicited events (arm B only; any grade):
    - solicited local injection site reactions: pain, itching, redness/blistering
    - solicited general reactions: fever, muscle aches, generalised itching, joint aches, headache, nausea, general discomfort
- SAEs, overall and by type of SAE (fatal, life-threatening, hospitalisation, disability, other).
- Lab abnormalities.
- Tabulation of adverse events of all grades

## 4.9 OTHER ENDPOINTS

The following variables will be analysed as for continuous outcomes as specified above:

- CD4 (cells/$mm^3$), CD4:CD8 ratio
- Quality of Life: summary score

In addition, the following will be presented over scheduled visits:

- Suppressed HIV-RNA: number (%)
- QTc interval ≤440ms: number (%)
- eGFR<60 ml/min/1.73$m^2$: number (%)
- Platelet count ≥150x$10^9$/L: number (%)
- Low haemoglobin (Hb≥12g/dL for males, ≥11g/dL for females): number (%)
- ALT <5xULN
- Merck p24 assay: To be specified.

# 5 REFERENCES

[1] Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, 3rd, Feldman HI, et al. A new equation to estimate glomerular filtration rate. Ann Intern Med. 2009;150(9):604-12.]

[2] Hancock G, Yang H, Yorke E, Wainwright E, Bourne V, Frisbee A, et al. (2015) Identification of Effective Subdominant Anti-HIV-1 CD8+ T Cells Within Entire Post-infection and Post-vaccination Immune Responses. PLoS Pathog 11(2): e1004658.