THE LANCET
HIV

Supplementary appendix 2

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Kekitiinwa AR, Jennings A, Bwakura-Dangarembizi M, et al. Efficacy and safety of short-cycle, dolutegravir-based antiretroviral therapy in adolescents living with HIV (BREATHER Plus): a multicentre, open-label, randomised, parallel, two-arm, non-inferiority trial. *Lancet HIV* 2026; published online April 7. https://doi.org/10.1016/S2352-3018(25)00326-1.

BREATHER Plus

A randomised open-label 2-arm, 96-week trial evaluating the efficacy, safety and acceptability of short cycle (five days on, two days off) dolutegravir/tenofovir-based triple antiretroviral therapy (ART) compared to daily dolutegravir/tenofovir-based triple ART in virologically suppressed HIV-infected adolescents aged 12 to 19 years of age in sub-Saharan Africa

ISRCTN: ISRCTN5058577

STATISTICAL ANALYSIS PLAN

Version 2.0

SAP version and approvals

Version Number and Date: 2.0 (17 June 2024) Supersedes version: 1.4 (13 June 2024) **In relation to protocol version:** 3.0			
Author	**Position**	**Signature**	**Date**
Angus Jennings	Delegated Statistician, MRC CTU at UCL		
Reviewed by			
Dr Deborah Ford	Trial Statistician (Senior Statistician), MRC CTU at UCL		
Professor Rodolphe Thiébaut	Independent Senior Statistician (reviewer), University of Bordeaux		
Approved by			
A/Professor Adeodata Kekitiinwa-Rukyalekere	Trial Chief Investigator, Baylor College of Medicine Children's Foundation-Uganda		
Professor Sarah Pett	Project Lead at MRC CTU at UCL, MRC CTU at UCL		

SAP revision history

Version	Date	Author	Comments
0.1	18 March 2020	Deborah Ford	Brief analysis plan described in protocol version 2.0.
0.2	03 December 2021	Ellen White	First draft
0.3	07 December 2021	Ellen White	Incorporating feedback on D3 SAP
0.4	08 December 2021	Ellen White	Incorporating feedback from Debbie Ford
IDMC Meeting (Discussion of protocol and SAP prior to enrolment opening) - 16 December 2021			
0.5	11 February 2022	Ellen White	Incorporating details on modified snapshot algorithm and non-inferiority margin in line with D3 SAP
0.6	25 February 2022	Ellen White	Incorporated feedback from Debbie Ford and Man Chan on modified snapshot algorithm
0.7	03 March 2022	Ellen White	Incorporated feedback from Debbie Ford and Sarah Pett on modified snapshot algorithm Incorporated additional non-inferiority margin information from Matteo Quartagno
0.8	07 March 2022	Ellen White	Incorporated feedback from Debbie Ford and Sarah Pett
0.9	04 April 2022	Ellen White	Incorporate feedback from Rodolphe Thiébaut on D3 SAP and at first IDMC meeting. Updated resistance and neuropsychiatric analysis sections.
1.0	06 April 2022	Ellen White	Up-versioned SAP
Pilot Study IDMC Review (via email) – 23 December 2022			
IDMC Meeting – 14 February 2023			
IDMC Meeting – 04 October 2024			
1.1	22 May 2024	Angus Jennings/Debbie Ford	Full specification of the MI approach to be used Include switch SCT-CT for viral rebound in primary endpoint SAE status when site and clinical reviewer disagree clarified Clarified that models will only adjust for randomisation stratification factors assuming there are sufficient participants in strata levels Safety analyses excluding spontaneous and induced abortions included
1.2	30 May 2024	Angus Jennings	Incorporated feedback from Debbie Ford Updated censoring dates from W102 to 'W102 – 1'
1.3	06 June 2024	Angus Jennings	Finalising Primary Analysis sensitivity analyses

1.4	13 June 2024	Angus Jennings	Modified section on Sensitivity Analysis (10.9) following Ian White's review of D3 SAP and amended one censoring date to account for any clinic visit with no VL measurement.
2.0	17 June 2024	Angus Jennings	Up-versioned SAP
IDMC Meeting – 27 June 2024			

Table of Contents

1. IMPORTANCE AND RATIONALE

The BREATHER PLUS trial will evaluate the virological efficacy, safety, acceptability and Quality of Life of DTG-based Short-cycle Therapy (SCT) with weekends off compared with Continuous Therapy (CT) with a DTG-based ART regimen. The backbone drugs will consist of tenofovir either as the TAF or TDF formulations partnered with either 3TC or FTC. Importantly for generalisability to low- and middle-income settings, the trial will be conducted using standard-of-care real-time viral load monitoring as recommended by the World Health Organization (currently six-monthly and annual in the adolescent population in sub-Saharan Africa); with additional plasma samples taken for safety monitoring by the Independent Data Monitoring Committee (IDMC) but not returned to doctors/patients.

2. TRIAL OBJECTIVES AND HYPOTHESIS

The overall aim of the BREATHER Plus trial is to assess whether DTG-based SCT with a tenofovir and lamivudine/emtricitabine backbone will provide non-inferior sustained virological suppression compared to continuous dolutegravir-based ART with a tenofovir and lamivudine/emtricitabine backbone.

2.1 PRIMARY OBJECTIVE

To assess whether DTG-based SCT with a tenofovir and lamivudine/emtricitabine backbone is non-inferior to continuous dolutegravir-based ART with a tenofovir and lamivudine/emtricitabine backbone in terms of virological suppression.

2.2 HYPOTHESIS

Dolutegravir-based SCT with a tenofovir and lamivudine/emtricitabine backbone will provide non-inferior sustained virological suppression compared to continuous dolutegravir-based ART with a tenofovir and lamivudine/emtricitabine backbone over 96 weeks.

3. ESTIMAND FOR PRIMARY ENDPOINT

Treatments	The comparison is between the SCT group and the CT group (control), as described in section 4.
Population	The population of interest is HIV-1 infected adolescents aged 12 to 19 years in Kenya, South Africa, Uganda, and Zimbabwe that meet the inclusion/exclusion criteria as defined in section 4.3.
Endpoint	Proportion of children with confirmed viral rebound, defined as the first of 2 consecutive HIV-1 RNA≥50 c/mL at any time up to the 96-week assessment (as defined in section 5.1)
Population-level summary measure	Difference in proportions (SCT - CT)
Intercurrent events	
Any treatment modification including: • change in any ART component; • ART dose modification; • ART discontinuation; • Return to continuous ART in the SCT group	Treatment policy

Missed doses of treatment	Treatment policy
Died	Hypothetical

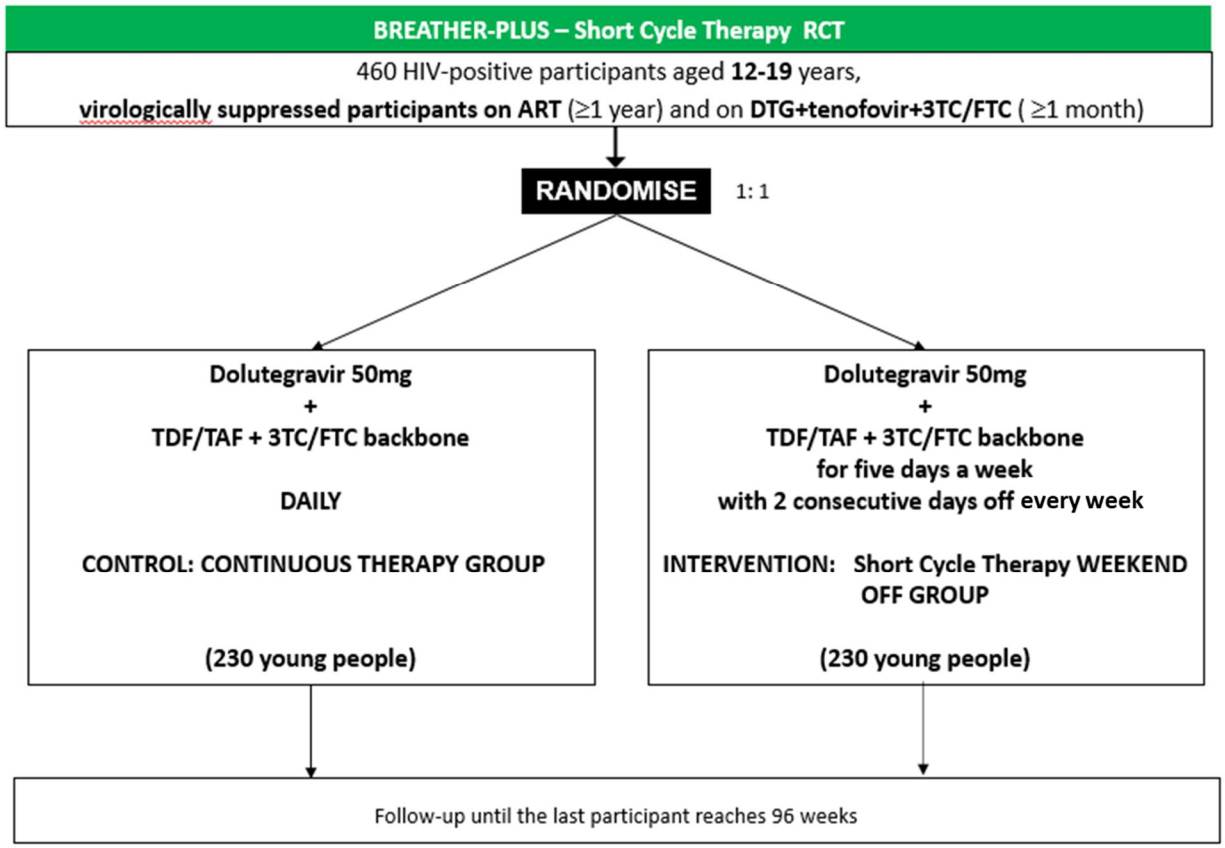
4. TRIAL DESIGN

4.1 STUDY DESIGN

The BREATHER Plus trial is an open-label, randomised (1:1), multicentre, non-inferiority trial in HIV-infected, non-pregnant, non-breastfeeding adolescents aged 12 to 19 years of age, virologically-suppressed for at least one year, without any history of treatment failure, on 3-drug combination antiretroviral (ART) consisting of dolutegravir with a 2-drug NRTI backbone consisting of tenofovir and lamivudine/emtricitabine for at least 1 month. All participants will be recruited in sub-Saharan Africa.[1]

A total of 460 adolescents will be randomised to DTG-based short-cycle therapy with weekends off (SCT) or continuous therapy with a DTG-based ART regimen (CT) (230 in each group). Randomisation will be a permuted block-randomisation approach, stratified by centre and mode of infection (horizontal or vertical). Visits are at screening, week 0 (randomisation), 4 (SCT only), 8, 16, 24, 32, 40, 48, 60, 72, 84, 96, and then every 12 weeks until the last participant reaches their 96 week visit.

Figure 4.1 Trial Entry, Randomisation and Treatment



Visits at screening, week 0 (randomisation), 4Ω, 8, 16, 24, 32, 40, 48, 60, 72, 84, 96, and then every 12 weeks until the last participant reaches their 96 week visit.

ΩSCT group only

[1] Full details of trial documents (including the Protocol) are stored in the BREATHER Plus eTMF

4.2 STUDY POPULATION

HIV-1 infected adolescents aged 12 to 19 years will be recruited from sites in Kenya, South Africa, Uganda, and Zimbabwe.

4.3 SELECTION OF PARTICIPANTS

There will be **no exceptions** to eligibility requirements at the time of randomisation. Questions about eligibility criteria must be addressed prior to attempting to randomise the participant. The eligibility criteria are the standards used to ensure that only medically appropriate patients are considered for this study. Patients not meeting the criteria should not join the study. For the safety of the participants, as well as to ensure that the results of this study can be useful for making treatment decisions regarding other patients with similar diseases, it is important that no exceptions be made to these criteria for admission to the trial.

Participants will be considered eligible for enrolment in this trial if they fulfil all the inclusion criteria and none of the exclusion criteria as defined below.

4.3.1 Participant Inclusion Criteria

1. HIV-1-infected
2. Aged 12 to 19 years
3. Aware of HIV status
4. On ART for ≥1 year, with no previous regimen change for treatment failure
5. On ART consisting of DTG, tenofovir and lamivudine/emtricitabine for ≥1 month prior to screening
6. Virologically suppressed with all HIV-1 RNA viral loads <50 c/mL[a] in the last 12 months up to and including screening. Additionally there must be one result <50 c/mL[a] at least 12 months prior to screening and the viral load at trial screening must be <50 c/mL
7. Girls who are sexually active must be willing to adhere to highly effective methods of contraception[b]
8. Written informed consent provided by participant (if aged 18 to 19 years) and/or carer/legal guardian (if participant aged 12 to 17 years) as appropriate
9. Written informed assent in participants aged 12 to 17 years

4.3.2 Participant Exclusion Criteria

1. Females who are pregnant or breastfeeding
2. Females who plan to become pregnant during the trial follow-up or are unwilling to use a highly effective method of contraception[b] for the duration of the trial if sexually active
3. Moderate or High risk score on the Columbia-Suicide Severity Rating Scale
4. On treatment for any active TB
5. Contraindication to continued receipt of dolutegravir or any formulation of tenofovir, lamivudine/emtricitabine
6. Underlying medical condition that in the opinion of the Investigator precludes participation
7. Previous randomisation in the LATA trial

[a]If a historic viral load is from a diluted sample (maximum dilution 1:5), and below lower limit of quantification (LLQ), a calculated VL<100 copies/mL is allowed; if the viral load in the diluted sample is equal to the LLQ, the calculated VL should be below 50 copies/mL. If there are any viral loads measured on dried blood spots since the most recent viral load on plasma more than 12 months ago these must be below the LLQ for the assay used. The screening sample viral load must always be <50 c/mL and cannot be done using a dry blood spot. [Protocol v3.0; amended between protocols for clarity only]

[b] Highly effective contraception are injectable, implantable, oral and intrauterine contraceptives which have an expected failure rate <1% per year

5. OUTCOME MEASURES

5.1 PRIMARY OUTCOME

The proportion of participants with confirmed viral rebound, defined as 2 consecutive plasma HIV-RNA ≥50 copies/mL at any time up to the 96-week assessment.

5.2 SECONDARY OUTCOMES

5.2.1 Efficacy

(i) Proportion of participants with HIV-RNA ≥50 copies/mL at 48 and 96 weeks using a modified FDA snapshot algorithm
(ii) The proportion of participants with HIV-RNA ≥1000 copies/mL (confirmed) by week 96
(iii) The number and type of HIV mutations at confirmed virological rebound
(iv) HIV-RNA <50 copies/mL and no switch to second-line ART for treatment failure at 24, 48, 72[2] and 96 weeks

5.2.2 Safety

(i) Change in toxicity profile including change in metabolic parameters (lipids, HbA1c, phosphate), renal function (eGFR) from baseline to 96 weeks; change in anthropometric measures from baseline to 48 and 96 weeks
(ii) Time to any new or recurrent WHO stage 3 or WHO stage 4 event or death
(iii) Incidence of serious, grade 3, 4 and 5, and treatment-modifying (of any grade) adverse events
(iv) The proportion of participants with any change from baseline ART regimen
(v) Change in CD4+ and CD8+ T-cell count from baseline to 48 and 96 weeks

5.2.3 Patient-reported outcomes

(i) Adherence, acceptability, wellbeing and including neuropsychiatric problems (e.g. depression, anxiety and sleep disturbance)
(ii) Healthcare resource utilisation (a sub-study outcome)
(iii) Health-related quality-of-life (a sub-study outcome)

6. SAMPLE SIZE CALCULATIONS

Non-inferiority of SCT will be assessed by the difference between the SCT group and the CT group in the estimated proportion of participants with viral rebound (defined as the first of two consecutive HIV-1 RNA ≥50 c/mL) by week 96.

The BREATHER Plus trial was designed with a fixed non-inferiority margin of 10%. At the design stage, it was estimated that a total of 460 participants (230 per arm) would provide 90% power to exclude a non-inferiority margin of 10% for the difference in the proportion of participants reaching the primary endpoint assuming 11% have had confirmed viral rebound (2 consecutive HIV RNA ≥ 50 c/mL) by 96 weeks in both arms, 10% loss to follow-up and a two-sided α of 0.05.

[2]Version 2.0 of the protocol stated that the visit schedule would be 8-weekly visits throughout the trial (week 0, 4 [SCT only], 8, 16, 24, 32, 40, 48, 56, 64, 72, 80, 88, 96 and then every 8 weeks), with a secondary outcome of HIV-RNA <50 c/mL and no switch to second-line ART for treatment failure defined at week 64. However, under protocol version 3.0, visits will be 12-weekly in year 2 (week 0, 4 [SCT only], 8, 16, 24, 32, 40, 48, 60, 72, 84, 96 and then every 12 weeks), where visits will no longer be conducted at week 64.

6.1 THE SAFE NON-INFERIORITY FRONTIER

There is uncertainty around the expected control event risk and the conservativeness of the assumed values. In order to protect against unexpected control event risks, methods based on non-inferiority frontiers will be implemented [1]. A non-inferiority frontier is a curve that defines the appropriate non-inferiority margin for each value of the control event risk. Most trials assume a fixed risk difference frontier, i.e. they are designed so that the non-inferiority margin remains fixed whatever the control event risk is. However, this has implications: if the control event risk turns out to be much lower than assumed, then a 10% non-inferiority margin might be considered too large, making results difficult to interpret; alternatively, if the control event risk turns out to be much larger than assumed, power would be lost if the same non-inferiority margin were used.

Given that the assumed confirmed viral rebound risk in CT arm ("control event risk") for the sample size calculation was conservative (i.e. the event risk was assumed on the high side, given the literature), it is important to protect against lower-than-expected risks (i.e. substantially lower than 11%), in order to preserve interpretability of results. A control event risk much larger than 11% is not considered likely in this trial. We will use the Smooth Away From Expected (SAFE) non-inferiority frontier for the main analysis, as shown in Figure 6.1. Accordingly, the non-inferiority margin to be used will depend on the observed confirmed viral rebound risk in the CT arm; should the confirmed viral rebound risk in the CT arm be substantially lower than 11% by 96 weeks, i.e. less than 9%, we will modify the non-inferiority margin (currently set at 10%). For example, if the confirmed viral rebound risk in CT arm was 5%, the non-inferiority margin would be modified to 8%; in the unlikely event that it was as low as 1%, the non-inferiority margin would be modified to 5%.

Simply using observed data to change the non-inferiority margin would lead to inflation of type 1 error. For this reason, if the non-inferiority margin is being changed, the significance level will be modified to control the type 1 error to ≤3%. Provided that the observed confirmed viral rebound risk in the CT arm is not lower than 9%, a 95% two-sided confidence interval will be computed for the difference in confirmed viral rebound between SCT and CT arms and a 10% non-inferiority margin will be used. If the observed confirmed viral rebound risk in the CT arm is less than 9%, a 99% two-sided confidence interval will be computed for the difference in confirmed viral rebound between SCT and SCT arms; the non-inferiority margin will depend on the observed confirmed viral rebound risk as shown in Figure 6.1 and Table 6.1. If the upper bound of the respective CI is no higher than the selected non-inferiority margin, then the null hypothesis will be rejected and SCT will be declared non-inferior to CT.

Different non-inferiority frontiers imply different null hypotheses, and hence different powers to reject the null under the alternative hypothesis. Table 6.1 shows the power of the trial if the SAFE frontier is used, depending on the true control event risk. The power for the expected event risk of 11% remains high at 87.5% and is always above 80% for lower control event risks. Power for larger event risks decreases, similarly to how it would change in a standard non-inferiority trial with fixed margin. Table 6.1 reports type 1 error rates as well, which remain close to 2.5% whatever the (unknown) true control event risk is. The probability of wrongly changing the margin under the null hypothesis is 15%.

Figure 6.1 Choice of non-inferiority margin based on observed confirmed viral rebound risk using the Smooth Away From Expected (SAFE) frontier

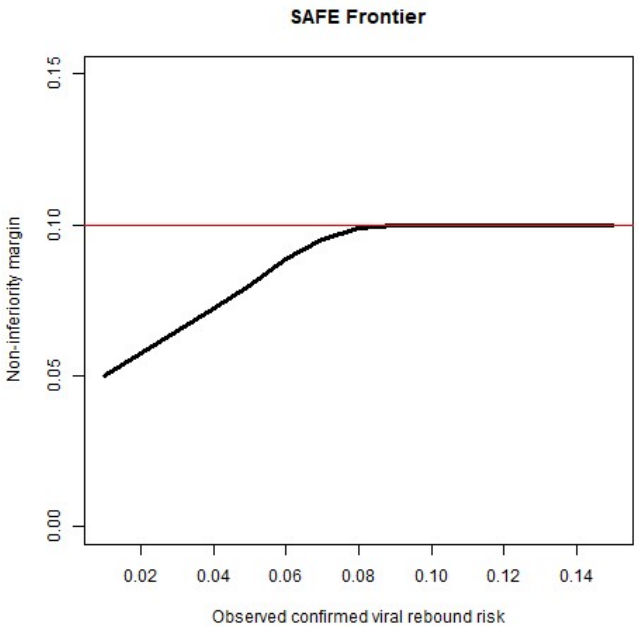


Table 6.1 Choice of non-inferiority margin and significance level based on observed confirmed viral rebound risk using the Smooth Away From Expected (SAFE) frontier

Observed confirmed viral rebound risk (P0)*	1%	2%	3%	4%	5%	6%	7%	8%	9%	10%	11%	12%	13%	14%	15%
NI Margin	5.0%	5.8%	6.5%	7.3%	8.0%	8.9%	9.5%	9.9%	10.0%	10.0%	**10.0%**	10.0%	10.0%	10.0%	10.0%
Significance level	0.5%	0.5%	0.5%	0.5%	0.5%	0.5%	0.5%	0.5%	2.5%	2.5%	**2.5%**	2.50%	2.50%	2.50%	2.50%
Power	95.8%	89.1%	84.4%	81.4%	80.4%	81.1%	82.2%	83.3%	84.6%	85.0%	**86.3%**	86.6%	85.7%	84.0%	82.0%
Type 1 error	2.50%	2.28%	2.36%	2.53%	2.67%	2.76%	2.86%	2.74%	2.69%	2.71%	**2.69%**	2.66%	2.64%	2.63%	2.61%
P (change margin)**	100%	100%	100%	99.9%	99.0%	95.0%	84.8%	67.8%	47.5%	29.0%	**15.5%**	12.1%	6.1%	2.8%	1.2%

The column in bold corresponds to the sample size calculation assumption made at the design stage.

*The choice of non-inferiority margin and significance level will depend on the observed confirmed viral rebound risk. The power, type 1 error and probability of changing margin depend on the true control event risk.

**The probability of changing the margin is the probability that, for a given true control event risk, the observed control event risk will be lower than 9%, hence leading to using a non-inferiority margin in the analysis different from the originally planned 10%.

7. ANALYSIS POPULATIONS FOR ANALYSIS

7.1 INTENTION-TO-TREAT POPULATION AND ANALYSIS

The intention-to-treat (ITT) population will consist of all randomised participants excluding those demonstrably randomised in error; where randomisation in error will be judged by the participant meeting a major violation of the eligibility criteria, including for example a participant being HIV negative, and will not depend on treatment allocation or post-randomisation follow-up. Major violation of the eligibility criteria will be determined by clinical decision, blind to treatment allocation.

The BREATHER Plus trial will be analysed as ITT. This will be the primary population used for the main analysis, which will use the randomised treatment allocation rather than actual treatment received. Primary inference will be based on the primary endpoint analysis of the ITT population. ITT was chosen over the per protocol analysis (described below) as it respects randomisation and ensures that the two treatment arms are, on average, comparable in terms of prognostic factors.*

7.2 PER PROTOCOL POPULATION AND ANALYSIS

The per protocol population will consist of all randomised participants excluding those who did not meet all the eligibility criteria and excluding any participants randomised in an incorrect stratum. Participants will also be excluded if they report taking <75% intended weekend breaks (SCT arm only; according to self-reported adherence) or <90% of time on ART (CT arm only; according to ART log) to earliest of 96 weeks or censoring date (see below).

Follow-up will be censored if a participant:

- had a break in any component of ART regimen for more than 7 days (SCT and CT arms);
- changed ART component for any reason, excluding changes between TDF and TAF and 3TC and FTC, and vice versa (SCT and CT arms);
- changed to continuous therapy for reasons other than confirmed viral rebound (SCT arm only)**.

A per protocol analysis of the primary endpoint will be done in the per protocol population.

* Some researchers consider that, in non-inferiority trials, a per protocol analysis is more appropriate than the ITT analysis because of the belief that under the null hypothesis of inferiority, the ITT analysis is anti-conservative in the presence of deviations from the protocol and missing data. However, simulation studies have demonstrated that this is not always the case and that the anticonservatism or otherwise depend on a number of factors including the nature of protocol deviations and the reasons for missing data (see for example [2, 3]). A per protocol analysis violates the integrity of the randomisation process. Consequently, estimates of the treatment difference obtained via per protocol analysis reflect a combination of a true treatment effect and subset selection bias and it is impossible to disentangle the two. These considerations which are widely accepted in superiority trials apply equally well to non-inferiority designs (see for example [4]).

**Note this would include changing from SCT to CT for pregnancy or while receiving TB treatment. Switch from SCT to CT for viral rebound following a single HIV-1 RNA ≥50 c/mL (where confirmatory measure was not performed before change in strategy) is considered "confirmed viral rebound" hence is not a reason for censoring (see section 10.8).

8. STRATA AND COVARIATES

8.1 STRATIFICATION VARIABLES

Randomisation will be stratified by:

- Centre
- Mode of infection (horizontal or vertical)[3]

[3] Vertical transmission defined as acquiring HIV through mother to child transmission (MTCT); horizontal transmission defined as acquiring HIV through other routes (sexual contact, blood product, unknown or other).

9. DATA HANDLING AND DERIVATION

9.1 DEFINITION OF BASELINE

Baseline is defined as the date of randomisation (day 0). For all variables, the week 0 measurement is defined as the latest single measurement up to 42 days prior to (and including) date of randomisation.

9.2 DEFINITION OF FOLLOW-UP TIME

Time will be measured from randomisation (day 0). For analysis of the primary endpoint, follow-up will be to end of week 96 window [week 102 date – 1 day]. For safety and occurrence of new or recurrent WHO 3 or WHO 4 event or death analysis, the primary analysis will include all follow-up to end of trial visit. If participants are censored earlier due to loss to follow-up or withdrawal of consent, it will be assumed that such censoring is independent of the outcome. For interim analyses, the date of data freeze will be used. For analyses by scheduled visits, only time-points where there are at least 10 participants in each group with a measurement will be included.

9.3 DEFINITION OF NOMINAL WEEK FOR CLINICAL AND LABORATORY MEASUREMENTS

Laboratory measurements, and other clinical parameters (e.g. weight), at any nominal week are defined as those taken nearest to the nominal week within equally spaced windows according to the protocol visit schedule (note that the visit schedule changes frequency). The midpoint between two scheduled visit weeks should be taken as belonging to the latter window (see below). Where there are two values within one of these equally spaced windows, but both equidistant from the nominal visit week, the later value will be used.

Week 4 (SCT only*):	2 ≤ week < 6 [between 2 & 6 weeks, including 2 but excluding 6]
Week 8*:	4 ≤ week < 12
Week 16:	12 ≤ week < 20
Week 24:	20 ≤ week < 28
Week 32:	28 ≤ week < 36
Week 40:	36 ≤ week < 44
Week 48:	44 ≤ week < 54
Week 60:	54 ≤ week < 66
Week 72:	66 ≤ week < 78
Week 84:	78 ≤ week < 90
Week 96:	90 ≤ week < 102

For all the subsequent visit weeks beyond week 96 (i.e. every 12 week follow-up visits), the same rule of +/- 6 weeks defining intervals will be applied as described above.

For laboratory measurements or assessments taken less frequently in line with the protocol schedule, the nominal week windows will be widened according to the above principle, as determined by scheduled visits and midpoint between scheduled visits. For measurements taken every 48 weeks, a rule of +/- 24 weeks defining intervals will be applied; measurements taken every 24 weeks, a rule of +/- 12 weeks will be applied; while measurements taken every 16 weeks, a rule of +/- 8 weeks will be applied.

*Week 4 visits will be conducted in the pilot and SCT arm only and will provide descriptive data at this timepoint. Week 8 is designed as the first comparative timepoint between trial arms, therefore week 4 measurements taken in SCT arm may contribute to cross-sectional week 8 analyses if they fall in the week 8 window and there is no nearer measurement. If anyone in either arm in the pilot, meets the primary endpoint at ≤4 weeks, they will be included as a failure at week 8.

9.4 HANDLING MISSING DATA

Complete case analysis will be performed for all analyses except for analyses using HIV-1 RNA and immunology data as described below.

9.4.1 HIV-1 RNA data

For analysis of the primary endpoint (section 10.8) and other virological outcomes (section 10.12.1), except for the FDA snapshot analysis, multiple imputation will be applied if either of the following criterium is met:

- 5% of all HIV-1 RNA measurements at scheduled visits are missing

Or

- 10% of confirmatory HIV-1 RNA measurements are missing.

If the criterium for imputation is met, missing HIV-1 RNA measurements at scheduled visits will be multiply imputed based on the procedure outlined in Appendix 1

9.4.2 Immunology data

Missing immunology data imputation will be performed as follows (following confirmation with participating labs):

- Immunology (CD4%, CD4+ count and total lymphocyte): If one of the three measurements is missing, the standard formula [CD4+ count = CD4% x Lymphocyte count/100] will be used to calculate the missing result using the remaining two available values. The corresponding formula is also used for CD8+.

9.5 HANDLING OF VIRAL LOAD DATA

The primary endpoint requires confirmation of HIV-1 RNA measurements ≥50 c/mL at any time up to end of week 96 window (week 102 date – 1 day), with the first of 2 consecutive HIV-1 RNA≥50 c/mL strictly measured prior to the end of week 96 window.

For real-time HIV-1 RNA (results obtained from "immediate" testing of samples taken at scheduled visits, which are entered directly onto the database by sites, as opposed to retrospective testing of stored samples), participants with HIV-1 RNA≥50 c/mL will be recalled within the respective visit week window to confirm their HIV-1 RNA result. EDTA-plasma samples will be stored at weeks 0, 4 (SCT only), 8, 16, 24, 32, 40, 48 and then 12-weekly from protocol version 3.0 for batched retrospective viral load testing where real time HIV-1 RNA results are not available (see Appendix 2 for proposed rules for requesting retrospective HIV-1 RNA testing from stored plasma samples and Section 10.8 for details of use of HIV-1 RNA data results for primary endpoint analysis).

The date of the first of two consecutive HIV-1 RNA measurements ≥50 c/mL will be used for calculating the time to the first confirmed measurement ≥50 c/mL (primary endpoint rebound date). For the cross-sectional analysis at weeks 24, 48, 72 and 96, where multiple measurements fall within the same visit window, the closest measurement to scheduled visit will be used as described in Section 9.3.

9.6 HANDLING ADVERSE EVENT DATA

Serious Adverse Events (SAEs) and notable events are reported to the CTU within 1 working day of the site becoming aware of the event. All SAEs, notable events, new or recurrent WHO 3 or 4 events, ART modifying events of any grade, clinical and clinically significant laboratory events of grade ≥3 during the trial will be reviewed by MRC CTU clinical reviewers. Asymptomatic laboratory events (that are not clinically significant) of grade ≥3 close in calendar time (±28 days) to clinical events will also be reviewed by the clinical reviewers to avoid double-reporting of events.

Clinical and laboratory adverse events will be graded according to the toxicity grading tables found in Appendix I of the protocol. Only new events or new recurrences of events that occur post-randomisation will be included in the analysis. This includes any event (clinical or laboratory) of grade ≥3 and ART modifying events (clinical or laboratory) of any grade. Ongoing events present at randomisation which had been reported prior to entry into the trial will only be reported as baseline data. For adverse event analysis, "Date of event diagnosis" on AE eCRF will be used as date of event. All deaths that occur after the date of randomisation up to end of trial will be included as BREATHER Plus events. The "date of death" will be used as the date of event for deaths. When calculating incidence rates, the time to the first event and recurrence will be used in the main comparison between the two study arms. For AEs where the site clinician and CTU clinical reviewer disagree on seriousness, the event will be reported as an SAE.

9.7 OTHER DEFINITIONS

- The WHO staging of HIV infection is defined according to diagnostic criteria for WHO stage 1, 2, 3 and 4 conditions [5] are described in the protocol (Appendix II).
- Clinical and laboratory adverse events will be classified by System Organ Class (SOC) according to version 25.0 of Medical Dictionary for Regulatory Activities (MedDRA).
- Gradings of clinical and laboratory adverse events (AEs) are defined according to the Adapted Division of AIDS (DAIDS) Table for Grading the Severity of Adult and Pediatric Adverse Events (Corrected Version 2.1 - July 2017), in the protocol (Appendix IV - Toxicity Gradings) [6]. Neutrophil grading is based on WHO guidelines [7] recognising the lower normal levels in African populations. For creatinine clearance DAIDS ranges of estimated creatinine clearance rather than comparison with the baseline will be used.
- New asymptomatic grade ≥3 laboratory events will be identified by detection of follow-up measurements of grade ≥3 starting from a baseline measurement of grade <3. The laboratory event will be classified as grade 3 or 4 according to the maximum grade achieved before resolution. Resolution is defined as return to a minimum of grade 2 or baseline grade.
- Pregnancies occurring during the trial are reported as notable events (protocol Section 7.2).

9.8 DATA QUALITY

General data quality is monitored within the OpenClinica database through real time generation of database queries, as described in the BREATHER Plus Data Management Plan. Additional data checks will be carried out at least monthly or as required by the statistician, depending on accrual, in order to ensure ongoing data quality. In addition to this, extensive data cleaning will take place before each IDMC meeting and final analysis.

10. STATISTICAL ANALYSES

10.1 GENERAL PRINCIPLES

The two arms of the trial will be compared as randomised according to the intention-to-treat principle. For the primary endpoint, a per protocol analysis will also be performed. Descriptive statistics will be reported overall and by randomised group, and percentages will be of non-missing values, with the number (%) of non-missing values given if data are not complete. Statistical tests will be 2-sided, estimates will be presented with a 2-sided 95% CI (or 99% as required per the primary analysis) and comparisons will adjust for the stratification factors (all combinations of centre and mode of infection, where sufficient participants are enrolled across strata). Appropriate transformations for all variables will be applied after inspection of the data.

Time-to-event outcomes will consider time from randomisation to the event date, using Kaplan-Meier estimation. Differences between treatment groups in time-to-event outcomes will be tested using adjusted Cox proportional hazard regression models (adjusted for stratification factors). All hazard ratios or estimates of the difference between groups will be presented with a 2-sided 95% CI and p-value.

Differences between randomised groups in binary outcome variables will be tested using Chi-squared tests or Fisher's exact tests, as appropriate. Fisher's exact test will be used if expected frequencies in one or more of the cells in 2x2 matrices is <5, otherwise Chi-square test will be used. Logistic regression models will be used for adjusted analyses (presented with 2-sided 95% CI).

Percentages will be reported to 0 decimal places, unless <0.5% when they will be given to one decimal place. P-values will be given to 3 significant figures. There will be no formal adjustment of α-values because of any interim analyses performed for IDMC meetings.

All analyses will be performed using the STATA software (updated and validated), unless otherwise specified.

10.2 ANALYSIS OF CONTINUOUS VARIABLES

For the analysis of continuous variables, the mean change from baseline over follow-up will be calculated using linear mixed models (unstructured covariance) with random intercept for participants and fixed effects for randomised group and visit weeks, including interaction between randomised group and visit weeks, adjusting for baseline. Adjusted analyses will be performed adjusting for stratification factors in addition to baseline.

For each of the scheduled follow-up visit weeks, the number (%) of participants with a measurement along with the following will be reported based on the above linear mixed models:

- Unadjusted (only adjusting for baseline) difference between arms: mean (SE), 95% CI, p-value.
- Adjusted (adjusting for baseline and stratification factors) difference between arms: mean (SE), 95% CI, p-value.
- Mean change from baseline, by arm (estimated from adjusted model, assuming baseline mean): mean (SE)

For each analysis, a graph showing the mean change from baseline across scheduled visits by randomised group will be presented including 95% CI.

Average treatment differences through follow-up will be estimated with 2-sided 95% confidence intervals by fitting linear mixed models with random intercept for participants and fixed effects for randomised group and visit weeks, adjusting for baseline measurement. The adjusted model will also adjust for stratification factors.

10.3 ANALYSIS OF ADVERSE EVENT DATA

Clinical and laboratory adverse events will be tabulated by MedDRA System Organ Class and Preferred Term (PT). Randomised groups will be compared in terms of time to first adverse event using Cox proportional hazard regression models calculating hazard ratios. Multiple events will be compared between randomised groups using Poisson regression models to calculate the incidence rate ratios of adverse events in the SCT vs CT arms adjusted for clustering within individuals. The 95% CI will be given along with p-value for the unadjusted and adjusted models. The adjusted model will be adjusted for the stratification factors. The total number of adverse events and number of participants with at least one adverse event will be computed. Adverse event rates (per 100 person years) will be calculated as the number of events/total person years at risk*100 (presented with 2-sided 95%CI).

10.4 ANALYSIS OF PARTICIPANT AND PARENT/CARER QUESTIONNAIRE DATA

Descriptive analysis of questionnaire data will be undertaken for carer and participant completed questionnaires separately. Participant and parent/carer reported binary responses will be compared between randomised groups over time using logistic mixed models with a random effect for intercept and fixed effects for treatment group, post-randomisation study visits and adjustment covariates [baseline answer and stratification factors]. The ordered categorical responses will be compared between randomised groups over time using ordered logistic mixed model with a random effect for intercept and fixed effects for treatment group, post-randomisation study visits and adjustment covariates [baseline answers and strata].

10.5 ANALYSIS DETAILS

10.5.1 Enrolment and eligibility

- **Enrolment over time:** plot by calendar month
- **Enrolment by arm:** number (%)
- **Enrolment by country and site:** number (%)
- **Enrolment by stratification factors:** number (%)
- **Numbers screened, randomised, not randomised including tabulation of reason for non-randomisation**
- **Details of minor and major ineligibility and number excluded from analysis:** number (%)
- **Trial CONSORT diagram:** details of the number of participants screened for eligibility, the number not randomised and reasons why, the number randomised to each arm, and the number in follow-up at week 96, and the number included in the primary analysis.

10.5.2 Baseline characteristics

Baseline characteristics will be summarised using descriptive statistics for all randomised participants. No statistical significance tests will be used to compare treatment arms.

10.5.2.1 Demographics

- **Sex at birth:** number (%) by sex (male, female)
- **Age (years):** number of participants and missing; mean, standard deviation (SD), median, interquartile range (IQR), range; number (%) in categories (12-<15, 15-<18, 18-<20 years).
- **Ethnic origin:** number (%) in categories (Asian, Black, Hispanic/Latino, Caucasian, other); number missing.

10.5.2.2 HIV and immune related parameters

- **Mode of infection:** number (%) in categories (mother to child, blood product, sexual transmission, other, unknown); number missing.
- **WHO staging for HIV:** number (%) in categories (1, 2, 3, 4); number missing.
- **CD4%:** number of participants and missing; mean, SD, median, IQR, range; number (%) in categories (<10%, 10-<15%, 15<20%, 20-<25%, 25-<30%, 30-<40%, ≥40%).
- **CD4+ absolute (cells/mm^3):** number of participants and missing; mean, SD, median, IQR, range; number (%) in categories (<350, 350-1000, 1000-1500, ≥1500).
- **CD8%:** number of participants and missing; mean, SD, median, IQR, range.
- **CD8+ absolute (cells/mm^3):** number of participants and missing; mean, SD, median, IQR, range.
- **Total Lymphocyte count (cells/mm^3):** number of participants and missing; mean, SD, median, IQR, range.
- **HIV-1 RNA:** number (%) in categories (<50 or ≥50 c/mL); number missing.

10.5.2.3 Growth parameters

- **Weight (kg):** number of participants and missing; mean, SD, median, IQR, range; number (%) in categories (<35kg, 35-<40kg and ≥40kg).
- **Height (cm), BMI (kg/m^2):** number of participants and missing; mean, SD, median, IQR, range.
- **Weight-for-age, height-for-age, BMI-for-age, z-score#:** number of participants and missing; mean, SD, median, IQR, range; number (%) in categories (<-3, -3 to <-2, -2 to <0, ≥0).
- **Waist (cm), hip (cm):** number of participants and missing; mean, SD, median, IQR, range.
- **Waist-hip ratio:** number of participants and missing; mean, SD, median, IQR, range.

#BMI, weight and height will also be analysed as BMI-for-age, weight-for-age and height-for-age z-scores. For BMI, weight and height, British 1990 Reference data (0–23 years), which covers the full age range of BREATHER Plus participants, will be used for standardisation.

10.5.2.4 Laboratory data

Lipid Results

- **Triglycerides (mmol/L):** number of participants and missing; mean, SD, median, IQR, range.
- **LDL Cholesterol (mmol/L):** number of participants and missing; mean, SD, median, IQR, range.
- **HDL Cholesterol (mmol/L):** number of participants and missing; mean, SD, median, IQR, range.
- **Total Cholesterol (mmol/L):** number of participants and missing; mean, SD, median, IQR, range.
- **Phosphate (mmol/L):** number of participants and missing; mean, SD, median, IQR, range.

Blood Biochemistry Results

- **Albumin (g/dL):** number of participants and missing; mean, SD, median, IQR, range.
- **Alanine transaminase (IU/L):** number of participants and missing; mean, SD, median, IQR, range.
- **Total Bilirubin (mg/dL):** number of participants and missing; mean, SD, median, IQR, range.
- **Creatinine (mg/dL):** number of participants and missing; mean, SD, median, IQR, range.
- **Creatinine clearance (eGFR)* (mL/min):** number of participants and missing; mean, SD, median, IQR, range.

*Calculation of eGFR will be performed using the Cockcroft-Gault formula

Haematology Results

- **Haemoglobin (g/dL):** number of participants and missing; mean, SD, median, IQR, range.
- **White blood cell count (cells/mm^3):** number of participants and missing; mean, SD, median, IQR, range.
- **Absolute lymphocyte count (cells/mm^3):** number of participants and missing; mean, SD, median, IQR, range.
- **Platelet count (cells/mm^3):** number of participants and missing; mean, SD, median, IQR, range.
- **Absolute neutrophil count (cells/mm^3):** number of participants and missing; mean, SD, median, IQR, range.
- **Glycosylated haemoglobin A1c (mmols/mol):** number of participants and missing; mean, SD, median, IQR, range.

10.5.2.5 Antiretroviral exposure pre-randomisation

- **Different ART drugs ever received, overall and by class:** median, range, summarised by categories (All classes, nucleoside/nucleotide reverse transcriptase inhibitor [NRTI], protease inhibitor [PI], non-nucleoside reverse transcriptase [NNRTI], integrase inhibitors [INSTI]).
- **Participants exposed to ART classes:** number (%) in categories (NRTI+PI only, NRTI+NNRTI only, NRTI+INSTI only, NRTI+PI+INSTI only, NRTI+NNRTI+INSTI only, NRTI+NNRTI+PI only, NRTI+NNRTI+PI+INSTI, DTG as a separate category).
- **Cumulative ART exposure in years, overall and by class:** median, range, summarised by categories (All classes, NRTI, PI, NNRTI, INSTI, DTG as a separate category)

10.5.2.6 Antiretroviral regimen at randomisation

- **ART regimen at randomisation by NRTI backbone:** number (%) in categories (TDF+3TC, TDF+FTC, TAF+3TC, TAF+FTC)

10.5.2.7 Questionnaire data

- Acceptability, neuropsychiatric symptoms, suicidal ideation and health-related quality of life questionnaire data will be summarised using frequency table and summary statistics.

10.6 FOLLOW UP AND AVAILABILITY OF HIV-1 RNA MEASUREMENTS

10.6.1 Follow up

Follow-up will be presented up to date of last follow-up or the date of data freeze for interim analysis, overall and by randomised group.

- **Number of participants randomised and seen after baseline.**
- **Weeks from randomisation to most recent study visit¥*:** number (%) in week categories (0, 1 to <4 weeks, 4 to <8 weeks, 8 to <16 weeks, 16 to <24 weeks, 24 to <32 weeks, 32 to <40 weeks, 40 to <48 weeks, 48 to <60 weeks, 60 to <72 weeks, 72 to <84 weeks, 84 to <96 weeks)
- **Median weeks from randomisation†:** median (IQR), range.
- **Weeks since last seen¥:** number (%) in week categories (Seen in last 0-8 weeks, seen in last 8-16 weeks, seen in last 16-24 weeks, seen in last 24-32 weeks, seen in last 32-40 weeks, seen in last 40-48 weeks). This will be reported at interim analyses.
- **Died or confirmed by site to have withdrawn consent:** number (%)
- **Potentially LTFU****: number (%)
- **LTFU/withdrawal:** number (%), timing and reason
- **Reached primary endpoint or seen at week 96: number (%)**
- **Scheduled visit attendance:** number (%) of scheduled study visits attended, overall and by assessment week.

* Week number is based on exact time in weeks rounded to the nearest week number given on CRF. This is not based on nominal windows around visit week.
† Using exact time rounded to the nearest week.
** Defined as either withdrawn consent or not seen for >24 weeks
¥ More week categories may be added as required.

10.6.2 Availability of HIV-1 RNA measurements

- **Availability at week X*:** number (%), available/missing

* Week number is determined by the window in which the visit date falls (see section 9.3). This means a participant may have a week X eCRF but if this does not fall in the window for week X and no other results fall in the window then they will be listed as missing.

10.7 TREATMENT ADHERENCE AND PROTOCOL DEVIATIONS

10.7.1 Use of antiretroviral therapy after randomisation

- **Follow-up regimen:** number (%) by anchor drug and NRTI backbone of regimen.
- **Tabulation of reason for switch to continuous therapy (SCT only):** number (%) switched to continuous therapy, time of switch and reasons for switch.
- **Tabulation of reasons for any ART changes:** number (%) on initial regimen and not on initial regimen. Number (%) changed from initial regimen and reasons for change or interruption. Initial regimen is defined as the regimen participants were on at time of randomisation: dolutegravir, TDF or TAF, and 3TC or FTC.
- **List of switches to continuous therapy (SCT only):** arm, trial number, date of switch, week, initial regimen, reason for switch
- **List of ART changes:** arm, trial number, date of change, week, initial regimen, regimen after change, reason for change in ART regimen.
- **List of reasons for treatment interruption:** arm, trial number, date of change, week, duration of treatment interruption in days, regimen interrupted and reason for treatment interruption.

- **Time to discontinuation of randomised treatment (weeks):** number of participants (%); mean, SD, median, IQR, range.

10.7.2 Protocol deviations

Protocol deviations are defined in the BREATHER Plus protocol deviations list and managed according to the BREATHER Plus Protocol Deviation Management Plan, stored in the BREATHER Plus eTMF. These two documents have been developed in line with the MRC CTU protocol deviations and violations standard operating procedures (SOP).

Protocol deviations: Number (%) by protocol deviation type (critical and major).

10.8 PRIMARY OUTCOME ANALYSIS BY WEEK 96

Participants will have an end of trial visit within ±6 weeks of the last recruited participant reaching 96 weeks follow-up (with return for confirmatory viral load if viral load is ≥50 c/mL at the end of trial visit). At analysis, a participant who does not attend their end of trial visit will be classified as "lost to follow-up" if they are not known to have died and the clinic has confirmed that they are unable to contact them.

The primary outcome measure is HIV-1 RNA ≥50 c/mL (confirmed on a separate sample) at any time up to end of week 96 window (week 102 – 1 day). Real-time HIV-1 RNA will be measured at weeks 48 and 96 (and 48-weekly thereafter). Participants with HIV-1 RNA ≥50 c/mL will be recalled for blood draw within the week 48 analysis window (44-54 weeks), week 96 analysis window (90-102 weeks), then 48 weekly (±6 weeks) and at the end of trial visit (+6 weeks) to confirm their HIV-1 RNA results. If a real time HIV-1 RNA is ≥50 c/mL at any other scheduled visits, participants will also be recalled for blood draw within the respective week window. A participant who has one HIV-1 RNA ≥50c/mL and switches ART for treatment failure or from SCT to CT for viral load rebound will be assumed to have met the primary outcome at the HIV-1 RNA ≥50c/mL prior to switch.

At visits where real time HIV-1 RNA are not measured, stored samples will be used for HIV-1 RNA measurement retrospectively (see Appendix 2 for proposed rules for requesting retrospective HIV-1 RNA testing from stored plasma samples; Appendix 3 for HIV-1 RNA data considerations/assumptions for analysis relating to presence of real-time and retrospective HIV-1 RNA results within same window and HIV-1 RNA results measured on diluted samples).

Date of first of 2 consecutive HIV-1 RNA ≥50 c/mL will be used as date of meeting the primary endpoint (virologic rebound). The study has been powered to examine the difference in proportion of children with confirmed viral rebound in SCT arm compared to the CT arm and therefore the conclusion regarding non-inferiority will be driven by this.

If the criterium for imputing missing results has not been met (section 9.4.1) and HIV-1 RNA measurement(s) are missing at a scheduled visit week(s) or at recall(s) for repeat:

- The participant will be assumed to be virologically suppressed between visits if the HIV-1 RNA result prior to missed visit/repeat is <50 c/mL.
- The participant will be assumed to be virologically suppressed from baseline until their first non-missing HIV-1 RNA result.
- The participant will be assumed to be virologically suppressed between visits if HIV-1 RNA result prior to missed visit/repeat is ≥50 c/mL and next available result is <50 c/mL.
- The participant will be considered to have met the primary endpoint for confirmed viral rebound if HIV-1 RNA result prior to missing visit/repeat is ≥50 c/mL and participant switches ART for treatment failure or from SCT to CT for or viral rebound before confirmatory HIV-1 RNA test. Date of confirmed viral rebound will be the date of HIV-1 RNA ≥50 c/mL prior to missed visit/repeat.
- The participant will be considered to have met the primary endpoint for confirmed viral rebound if HIV-1 RNA prior to missed visit/repeat is ≥50 c/mL and next available result is ≥50 c/mL. Date of confirmed viral rebound will be the date of HIV-1 RNA ≥50 c/mL prior to missed visit/repeat.

- The participant will be censored if the last available HIV-1 RNA is ≥50 c/mL, after which the participant is LTFU, withdrew or died prior to end of week 96 window. Participant will be censored at the last HIV-1 RNA<50 c/mL before last available HIV-1 RNA ≥50 c/mL (or at baseline if the only available viral load during trial follow-up was ≥50 c/mL).

For those participants who have not already reached the primary endpoint by their last visit prior to end of week 96 window [week 102 date – 1 day], we will determine their primary endpoint status as follows, provided that criterium for imputing missing results has not been met (section 9.4.1):

Last HIV-1 RNA *before* W102 date* (end of week 96 window)	**Last seen in clinic *on or after* W102 date** (end of week 96 window)	**Primary endpoint status**	**Primary endpoint date**
<50 c/mL	Yes	Censored	W102 date – 1 day
	No	Censored	Date of last HIV-1 RNA
≥50 c/mL§** (no repeat)	Yes & first HIV-1 RNA on/after W102 date <50 c/mL	Censored	W102 date – 1 day
	Yes & no HIV-1 RNA on/after W102 date	Censored	Date of last HIV-1 RNA<50c/mL before the last HIV-1 RNA≥50c/mL before W102 date
	Yes & first HIV-1 RNA on/after W102 date ≥50 c/mL	Reached endpoint	Date of last HIV-1 RNA ≥50 c/mL before W102 date
	No	Censored	Date of last HIV-1 RNA<50 c/mL before the last HIV-1 RNA≥50 c/mL before W102 date

***** The last HIV-1 RNA result can be at any scheduled visit after randomisation before end of week 96 window. Anyone with no viral load data after randomisation (or week 4 in the SCT arm, where HIV1-RNA<50c/mL at week 4) will be censored at baseline. For example, if the participant was LTFU or withdrew or died at week 48, their last available HIV-1 RNA may be at week 48. The table illustrates how we will determine the primary endpoint status by week 96 depending on the last HIV-1 RNA result (< or ≥50 c/mL) available before end of week 96 window and if participants are last seen on or after end of week 96 window with or without HIV-1 RNA result.

****** Last HIV-1 RNA≥50 c/mL available before end of week 96 window missing a repeat could be a real time measurement ≥50 c/mL missing a repeat; or, a measurement obtained from retrospective testing of stored sample, where repeat request is not applicable.

§ Treatment switch for failure or change from SCT to CT for viral rebound on or after an unconfirmed HIV-1 RNA≥50c/ml will be considered as confirmation of viral rebound; date of primary endpoint date will be date of last HIV-1 RNA≥50c/mL before W102 date.

If the criterium for imputation is met (9.4.1), missing HIV-1 RNA measurements will be multiply imputed as described in Appendix 1. For analysis of the primary endpoint by week 96, missing HIV-1 RNA measurements at any of the scheduled follow-up visits up to and including week 108 visit will be multiply imputed.

For those participants who have not already reached the primary endpoint by their last visit prior to end of week 96 window [week 102 date – 1 day], we will determine their primary endpoint status as follows, provided that the criterium for imputation has been met:

Last HIV-1 RNA (in W96 window¥) *before* W102 date (end of week 96-window)	**Last seen in clinic *on or after* W102 date** (end of week 96 window)	**Primary endpoint status**	**Primary endpoint date**
<50 c/mL	Yes	Censored	W102 date – 1 day
	No	Censored	Date of last HIV-1 RNA
≥50 c/mL (no repeat observed)**ǂ §**	Yes & W108 HIV-1 RNA<50 c/mL	Censored	W102 date – 1 day
	Yes & W108 HIV-1 RNA≥50 c/mL	Reached endpoint	Date of last HIV-1 RNA≥50 c/mL before W102 date
	Imputed W108 HIV-1 RNA	To be determined by imputed W108 HIV-1 RNA **(Reached endpoint if imputed HIV-1 RNA≥50 c/mL)**	To be determined by imputed W108 HIV-1 RNA **(Date of last HIV-1 RNA≥50 c/mL before W102 date if reached endpoint)**
		To be determined by imputed W108 HIV-1 RNA **(Censored if imputed HIV-1 RNA<50 c/mL)**	To be determined by imputed W108 HIV-1 RNA **(W102 date – 1 day if censored)**

¥ This table refers to last HIV-1 RNA in W96 window because it illustrates how we will determine the primary endpoint status by week 96 given that all missing results at previous scheduled visits will have been multiply imputed.
ǂ Last HIV-1 RNA≥50 c/mL before W102 date with no repeat observed could be, for example, a real time week 96 HIV-1 RNA measurement where repeat sample was lost and no further samples available.
§ Treatment switch for failure or change from SCT to CT for viral rebound on or after an unconfirmed HIV-1 RNA≥50c/ml will be considered as confirmation of viral rebound; date of primary endpoint date will be date of last HIV-1 RNA≥50c/mL before W102 date.

For the primary analysis, the two treatment groups (SCT and CT) will be compared in the intention-to-treat population. The comparison will be of the cumulative probability of confirmed viral rebound by week 96 (as defined above); multiply imputed HIV-1 RNA data will be used as described in section 9.4.1. To allow for censoring, the survival curve for each combination of strata and randomised group will be calculated using a Cox model adjusting for stratification factors (as appropriate where sufficient participants are enrolled across strata) and randomised group. The average cumulative failure function (1-survival curve) for each randomised group will be estimated by standardisation procedure [8] as a weighted average of the corresponding stratum-specific cumulative failure functions with weights equal to the prevalence of that stratum in the whole ITT population. The difference in the probability of virological rebound between the SCT and CT arms will be estimated by the average difference between the cumulative failure functions at week 96. A 2-sided bias-corrected 95% CI or 99% (Selected as per in section 6) for the difference in the probability of confirmed viral rebound by week 96 (SCT – CT) will be calculated using appropriate (bias-corrected) percentiles of the bootstrap estimates. The bootstrapping will sample 10,000 times and be stratified by stratification factors. SCT will be considered non-inferior to CT if the upper limit of the respective confidence interval of the difference SCT-CT is less than the selected non-inferiority margin (Selected as per in section 6). The 95% CI or 99% of the estimated probabilities of confirmed viral rebound by week 96 in each randomised group will also bias-corrected.

Kaplan-Meier plots of cumulative probability of confirmed viral rebound stratified by treatment allocation to week 96 will be produced, alongside plots of adjusted cumulative probability of confirmed viral rebound by treatment allocation based on the Cox model.

A per protocol analysis of the primary endpoint will be carried out in the per protocol population as described in section 7.2. Analysis will be done as described previously for primary outcome analysis this section. If criterium for multiple

imputation has been met (section 9.4.1), multiply imputed HIV-1 RNA data used for primary endpoint analysis will be used.

In addition, since there may be other reasons for return from SCT to CT (e.g. persistent viral blips or clinician concern about adherence), we will consider a failure endpoint comprising HIV-1 RNA ≥50 c/mL (confirmed on a separate sample) or switch off randomised treatment regimen or strategy (except for treatment/strategy change for pregnancy) at any time up to end of week 96 window.

10.9 SENSITIVITY ANALYSES FOR THE PRIMARY OUTCOME

The following sensitivity analyses will be performed:

I. Analysis accounting for non-proportional hazards

The assumption of proportional hazards in the Cox regression model will be assessed through visual consideration of treatment-stratified Kaplan-Meier plots and plots of observed $\ln(-\ln(S(t))$ (requiring parallel lines for the two treatment groups). Sensitivity of conclusions to the proportional hazards assumption will be assessed by fitting a flexible parametric model[9], allowing the treatment effect to vary over follow-up using a restricted cubic spline function. In the case that proportionality of hazards is violated and overall conclusions differ with respect to non-inferiority of SCT, then the flexible parametric model will be used for the primary analysis.

II. Analysis unadjusted for stratification factors

Analysis unadjusted for stratification factors will be done as described previously for primary outcome analysis in section 10.8. If criterium for multiple imputation has been met (section 9.4.1), multiply imputed HIV-1 RNA data used for primary endpoint analysis (section 10.8) will be used. The bootstrapping will sample 1,000 times rather than 10,000 times used for primary outcome analysis.

III. Analysis using non-imputed data

If the criterium for imputing missing HIV-1 RNA results has been met, analysis of non-imputed data will additionally be done as described in section 10.8. The bootstrapping will sample 1,000 times rather than 10,000 times used for primary outcome analysis.

10.10 SUBGROUP ANALYSIS FOR THE PRIMARY OUTCOME

Subgroup analysis for primary endpoint will be performed by randomisation stratification factors (centre and mode of transmission [vertical, horizontal]), where sufficient participants are enrolled and by age category at baseline [12-<15, 15-<18, 18-<20 years].

10.11 SECONDAY ANALYSIS OF PRIMARY ENDPOINT

Secondary analyses comparing time to confirmed viral rebound between arms will be performed using Cox proportional hazard regression models, adjusted and unadjusted for the stratification factors. The adjusted and unadjusted hazard ratios will be presented with a 2-sided 95% confidence interval.

10.12SECONDARY OUTCOME ANALYSES

10.12.1 Efficacy

10.12.1.1 Proportion of participants with HIV-1 RNA ≥50 c/mL at 48 and 96 weeks using a modified version of the FDA snapshot algorithm

The proportion of participants with HIV-1 RNA ≥50 c/mL at weeks 48 and 96 will be compared between arms using a modified version of the FDA snapshot algorithm (see Appendix 4) [10]. The estimated difference in proportion between SCT and CT arms will be computed with 95% CI by the Mantel-Haenszel weighted mean of proportions in each stratum.

10.12.1.2 The proportion of participants with HIV-RNA ≥1000 c/mL (confirmed) by week 96

The cumulative probability of confirmed viral rebound (defined as the first of two consecutive HIV-1 RNA ≥1000 c/mL) by week 96 will be estimated as described previously in section 10.8 (primary outcome analysis by week 96). If criterium for multiple imputation has been met (section 9.4.1), multiply imputed HIV-1 RNA data used for primary endpoint analysis (section 10.8) will be used. The bootstrapping will sample 1,000 times rather than 10,000 times used for primary outcome analysis by week 96.

10.12.1.3 The number and type of HIV mutations at confirmed viral rebound

In participants with confirmed HIV-1 RNA ≥50c/mL, resistance testing at failure will be performed on stored plasma samples. Activity of the background regimen will be defined according to the most current Stanford database algorithms at time of analysis. NRTI, PI, NNRTI and INSTI resistance will be defined according to the most current IAS-USA list of mutations at the time of analysis. Resistance mutations (overall and by ART class) detected at failure will be tabulated by number (%) of participants and type of mutation. The proportion of participants with resistance at failure will be compared between arms using the Chi-squared test or Fisher's exact test, as appropriate. 95% confidence interval for the difference in proportion will be provided. Logistic regression will be used for analyses adjusting for stratification factors.

10.12.1.4 Proportion of participants with HIV-1 RNA <50 c/mL and no switch to second-line ART for treatment failure at weeks 24, 48, 72[4] and 96 (cross-sectional analysis)

To evaluate the crude proportion difference between the arms at weeks 24, 48, 72 and 96, Fisher's exact test or Chi-squared test will be used to assess the difference between the SCT and CT arms in the proportion of participants with HIV-1 RNA <50 c/mL and no switch to second-line ART for treatment failure*. 95% confidence interval for the estimated difference in crude proportion of virological rebound at weeks 24, 48, 72 and 96 will be provided using normal approximation. Where appropriate, logistic regression will be used for analyses adjusting for stratification factors. If criterium for multiple imputation has been met (section 9.4.1), multiply imputed HIV-1 RNA data used for primary endpoint analysis (section 10.8) will be used.

*Switch for treatment failure as any change to third agent where reason for change is viral rebound or failure (clinical indication (i.e. what they say on ART log) and follows at least one real-time VL≥50 c/mL). Note that participants may meet the primary endpoint and delay or not switch to second-line ART.

[4]Version 2.0 of the protocol stated that the visit schedule would be 8-weekly visits throughout the trial (week 0, 4 [SCT only], 8, 16, 24, 32, 40, 48, 56, 64, 72, 80, 88, 96 and then every 8 weeks), with a secondary outcome of HIV-RNA <50 c/mL and no switch to second-line ART for treatment failure defined at week 64. However, under protocol version 3.0, visits will be 12-weekly in year 2 (week 0, 4 [SCT only], 8, 16, 24, 32, 40, 48, 60, 72, 84, 96 and then every 12 weeks), where visits will no longer be conducted at week 64.

10.12.2 Safety

10.12.2.1 Change in metabolic parameters from baseline to 96 weeks

Mean changes in lipids, HbA1c and phosphate from baseline will be reported by scheduled visit week and overall, over follow-up. Analysis of the mean change from baseline will be performed as described in section 10.2.

10.12.2.2 Change in renal function from baseline to 96 weeks

Mean changes in eGFR* from baseline will be reported by scheduled visit week and overall, over follow-up. Analysis of the mean change from baseline will be performed as described in section 10.2.

*Calculation of eGFR will be performed using the Cockcroft-Gault formula.

10.12.2.3 Change in anthropometric measures from baseline to 48 and 96 weeks

Mean changes in height, weight, BMI, height-for-age, weight-for-age, BMI-for-age z-scores, waist, hip, waist-hip ratio, from baseline will be reported by scheduled visit week and overall, over follow-up. Analysis of the mean change from baseline will be performed as described in section 10.2.

10.12.2.4 Time to any new or recurrent WHO 3 or WHO 4 events or death

Randomised groups will be compared in terms of time to first event using Cox proportional hazard regression models (adjusted for the stratification factors). Hazard ratios for first events will be presented with a 2-sided 95% confidence interval. Event rates will be calculated as the number of first events/total person years at risk*100 and will be reported per 100 person years including 95% confidence intervals.

10.12.2.5 Incidence of serious adverse events (SAEs)

Serious adverse events will be tabulated separately within the table as follows:

- **Total number of SAEs (number of participants)**
- **Number of SAEs (number of participants) by type of SAE:**
  - Fatal
  - Life-threatening
  - Hospitalisation
  - Persistent or significant disability/incapacity
  - Congenital anomaly/birth defect
  - Important medical condition that carries a real risk of one of the outcomes
- **Number of SAEs (number of participants) by:**
  - System Organ Class (SOC)
  - Preferred Term (PT) within System Organ Class

Hazard ratios, incidence rate ratios and event rates for serious adverse events will be calculated as described previously in section 10.3.
A secondary comparison of the incidence of SAEs between arms will also be made excluding all abortions (spontaneous and induced).

10.12.2.6 Incidence of Grade≥3 clinical and laboratory adverse events (AEs)

Clinical AEs are graded by the clinician reporting the event. Laboratory AEs of Grade≥3 will be graded using the limits defined in Appendix I of the protocol.

Clinical and laboratory AEs will be tabulated separately within the table as follows:

- **Total number of Grade≥3 AEs (number of participants)**

- **Number of Grade≥3 AEs (number of participants) by:**
  - System Organ Class
  - Preferred term within System Organ Class

Hazard ratios, incidence rate ratios and event rates for Grade≥3 AEs will be calculated as described previously in section 10.3.
A secondary comparison of the incidence of Grade≥3 AEs between arms will also be made excluding all abortions (spontaneous and induced).

10.12.2.7 Incidence of adverse events leading to discontinuation or modification of the treatment regimen

Discontinuation or modification of treatment regimen for AE is defined as change of any ART component for AE including dose increase due to use of concomitant medication required for treatment of the AE (e.g. increase in DTG dose due to starting rifampicin for TB treatment). Treatment interruption for >31 days for AE is considered to be discontinuation of treatment regimen for AE. In a secondary analysis, we will exclude AEs which only led to dose modification related to use of concomitant medication (e.g. increase in DTG dose due to starting rifampicin for TB).

AEs (clinical or laboratory) of any grade leading to discontinuation or modification of the treatment regimen will be tabulated as follows:

- **Total number of AEs (number of participants)**
- **Number of AEs (number of participants) by:**
  - System Organ Class
  - Preferred term within System Organ Class.

Hazard ratios, incidence rate ratios and event rates for adverse events leading to discontinuation or modification of the treatment regimen will be calculated as described previously in section 10.3.

10.12.2.8 Proportion with a change in baseline ART regimen

Change in baseline ART regimen will be defined as a change to any component of the baseline regimen (excluding switched between TDF and TAF and 3TC and FTC, and vice versa). Changes in baseline ART regimen during the trial will be tabulated by number (%) of participants. The proportion of participants with changes in baseline ART will be compared between arms using the Chi-squared test or Fisher's exact test, as appropriate. 95% confidence interval for the difference in proportion will be provided. Logistic regression will be used for analyses adjusting for stratification factors.

10.12.2.9 Change in CD4+ and CD8+ count from baseline to weeks 48 and 96

Mean changes in CD4+ and CD8+ from baseline will be reported by scheduled visit week and overall, over follow-up. Analysis of the mean change from baseline will be performed as described in section 10.2.

10.12.3 Patient-reported outcome measures (questionnaires)

10.12.3.1 Adherence, acceptability and wellbeing questionnaires

Descriptive analysis of the adherence and acceptability (HAT-QOL) questionnaires will be undertaken. The proportion of questionnaires where the participant reports missing any dose in the last week (ignoring weekend breaks allowed for in SCT arm) will be compared between randomised groups over time using logistic mixed models, as described in section 10.4.

10.12.3.2 Neuropsychiatric problems (mood survey questionnaire)

Descriptive analysis of the mood survey questionnaire (containing questions on depression, anxiety and sleep disturbance*) will be performed. The proportions reporting symptoms will be compared between randomised groups over time using logistic mixed models or ordered logistic mixed models, as appropriate, as described in section 10.4.

* Three longer questionnaires on depression (PHQ-9), anxiety (GAD-7) and sleep disturbance (based on Pittsburgh sleep questionnaire) will be completed only by participants in the neuropsychiatric toxicity sub-study and will be analysed separately.

10.12.3.3 Suicidal ideation and behaviour

Analysis of suicidal ideation and behaviour data will be performed as per guidance from the "Columbia–Suicide Severity Rating Scale Scoring and Data Analysis Guide" [11]. Analysis will be performed as described in section 10.4.

10.12.3.4 Quality of life (EQ-5D-Y) questionnaires

A descriptive analysis of each dimension of the EQ5D will be undertaken, as described in section 10.4.

10.12.4 Other outcomes

10.12.4.1 Changes in other laboratory measurements from baseline to week 96

Changes in other laboratory measurements from baseline will be reported by scheduled visit week and overall, over follow-up. Analysis of the mean change from baseline will be performed as described in section 10.2.

Analysis will include assessment of changes in:

- Total lymphocyte count (absolute and percentage).
- Haemoglobin, platelets, white cell count, neutrophil and lymphocyte counts, glycosylated haemoglobin A1c. Creatinine, albumin, total bilirubin, alanine transaminase (ALT).

10.12.4.2 MEMS Caps

For participants in the MEMS Cap sub-study, descriptive analyses will report MEMS Cap opening data by day of the week and trial arm.

10.13 TESTING MULTIPLE SECONDARY OUTCOMES

We will make no adjustment to p-values or confidence intervals to allow for testing multiple secondary outcomes. This study has a single primary outcome (confirmed viral rebound by week 96). The primary analysis is non-inferiority of SCT versus CT. If the confidence interval for the treatment effect (difference in proportion experiencing confirmed viral rebound by 96 weeks (SCT-CT)) lies below the selected non-inferiority margin (selected according to the SAFE non-inferiority frontier described in Section 6.1), then we will also test for superiority; because this is a closed test procedure there is no issue of multiplicity.

Secondary outcomes are divided in the protocol into efficacy outcomes, safety outcomes and other patient-reported outcomes. For safety outcomes it is appropriate to test each independently since it is important to identify any risks associated with SCT. The secondary efficacy outcomes are very closely related to the primary outcome. We will not adjust for multiple testing for these since they are correlated with the primary outcome (so standard adjustments are conservative). Resistance and immunological outcomes are considered exploratory and significance tests on these outcomes alone will not be used to conclude superiority.

We will report significance tests for differences between treatment arms for patient-reported outcomes, but if we have failed to demonstrate non-inferiority of SCT versus CT for the primary outcome, we will not use significance tests for these patient-reported outcomes to conclude superiority.

11. SUBSTUDIES

The following sub-studies will be conducted in subsets of participants:

i. Social science sub-study which will quantitatively and qualitatively assess adherence, acceptability and well-being among trial participants.

ii. Neuropsychiatric toxicity sub-study. Specific objectives include: To compare neuropsychiatric toxicities, including depression, suicidality, anxiety and sleep disturbance longitudinally between randomised groups; to test practical and feasible tools to identify and monitor mental health illness among adolescents in busy over-stretched HIV clinics.

iii. Health economics. Specific objectives are to assess the costs and cost-effectiveness of SCT compared to CT.

12. TIMING OF INTERIM ANALYSIS

12.1 PILOT STUDY

Among the first participants randomised into the trial, 15 participants randomised to SCT and 15 randomised to CT at sites participating in the pilot study will have viral load measurements at weeks 1, 2, and 3 (and a confirmatory viral load at week 4 following a single viral load above 50 c/mL at week 3) with these measures in the DTG SCT group being performed after weekends off treatment. Should any participants randomised to SCT in the pilot not take weekends off treatment, additional SCT participants will be recruited. The Independent Data Monitoring Committee (IDMC) will review these HIV RNA assessments after all participants in the pilot phase have completed this pilot. This will be to determine whether the SCT DTG group is safe to continue. Recruitment to the trial will continue during pilot follow-up and review of the pilot data.

12.2 INTERIM ANALYSES

The IDMC will meet to review unblinded data for randomised comparisons within 6 months and 12 months of the trial starting. Batch runs of viral loads will be planned prior to IDMC meetings and should include test results for visits up to 3-4 months prior to the IDMC meeting date; by the 12-month review meeting viral loads 8 and 16 weeks after enrolment should be available on ~140 participants, with 24-week viral loads on ~70 participants. Refer to the IDMC charter for guidance provided on stopping the trial early.

The frequency of future IDMC meetings will be at the discretion of the IDMC but are likely to be every 6-12 months.

13. TIMING OF FINAL ANALYSIS

Analysis of all outcomes will be completed within 6-12 weeks of end of trial defined in protocol: when all participants have attended their final study visit (including follow-up for VL HIV-RNA ≥50 c/mL), retrospective viral load testing is complete, and the database has been locked.

14. APPENDICES

14.1 APPENDIX 1 – HANDLING MISSING DATA WITH MULTIPLE IMPUTATION

This appendix describes the analysis that will be performed if either of the two criteria for multiple imputation listed in section 9.4 of the main statistical analysis plan are met.

14.1.1. GENERAL PROCEDURE

14.1.1.1 Handling missing log_{10}-transformed viral load measurements at baseline

Missing viral load (VL) data at baseline are expected to be minimal, and participants should be virologically suppressed at baseline. If there are missing log_{10}-transformed VL measurements at baseline, the missing values will be replaced with the overall mean of the observed log_{10}-transformed VL measurements. Mean imputation has been shown to be a valid approach for handling missing data in baseline covariates[12]. Baseline VL measurements known to be below the limit of detection (i.e. left-censored) will be handled using the procedure described in section 14.1.2. Note that baseline VL data do not contribute to the derivation of the primary endpoint and are therefore only used to inform imputation of follow-up VL data.

14.1.1.2 Imputing log_{10}-transformed viral load measurements post baseline

Multiple imputation of log_{10}-transformed VL measurements at 12 scheduled visits post-baseline (weeks 8, 16, 24, 32, 40, 48, 60, 72, 84, 96, 108, and 120) will be performed by multivariate imputation by chained equations (via mi impute chained in Stata)[13], based on the assumption of data being missing at random[14]. Week 120 visit window will be included in the imputation as an auxiliary variable (i.e. to inform imputation of week 108 visit window, and not to derive the primary endpoint), provided that there are at least 50% of participants with a VL measurement in that window.

Imputation will be performed separately in each randomised arm (via mi impute chained, by() in Stata) to allow for any interactions with randomised arms.

One log_{10}-transformed VL measurement per visit window will be included in the imputation model as a separate continuous variable (i.e. 12 incomplete variables to be imputed). For visit windows with more than 1 VL measurement, the choice of which measurement to include is described below.

Information on treatment changes will be summarised in the following two variables which will then be used to inform imputation of missing VL data:

- Before imputation, an overall binary indicator of whether each participant made a change to their originally randomised treatment (excluding changes between TDF and TAF or 3TC and FTC) or strategy (SCT to CT) (at any time after randomisation and before week 120), will be created; this indicator of treatment change takes value 1 if the participant made a treatment change, and 0 if the participant did not make a treatment change;

- Another variable containing time to treatment change will also be created. For participants who made a treatment change (i.e. for whom the binary indicator of treatment change takes value 1), this variable is defined as time (in weeks) from randomisation to the date of the treatment change. For participants who did not make a treatment change, this variable is replaced with time (in weeks) from randomisation to the scheduled visit date of their 120-week visit.

Each missing VL measurement per visit window will be imputed using predictive mean matching, using 11 nearest neighbours (via mi impute chained(pmm, knn(11))in Stata), conditional on the following:

- observed and imputed VL measurements from all other post-baseline visit windows;
- observed and mean imputed baseline log_{10}-transformed VL data;
- an overall binary indicator of whether the participants made a change to their originally randomised treatment (as defined above);

- a variable containing the time from randomisation to treatment change (as defined above); and
- site.

If the imputation procedure fails to converge in either arm (in more than 5% of any of the bootstrapped datasets, see section 14.1.3), the imputation models will be simplified for both arms (and for all bootstrapped datasets, see section 14.1.3), where each missing VL measurement per visit window will be imputed conditional on the following:

- observed and imputed VL measurements from ± 1 adjacent window (e.g. VL data at week 48 will be imputed using VL data at weeks 40 and 60, via the option omit() in mi impute chained in Stata);
- observed and mean imputed baseline VL data;
- an overall binary indicator of whether the participants made a change to their originally randomised treatment (as defined above);
- a variable containing the time from randomisation to treatment change (as defined above); and
- site.

The multivariate imputation by chained equations algorithm is an iterative procedure that cycles through the incomplete variables to be imputed a number of times before an imputed dataset is created, in order to ensure convergence of the conditional imputation models. Previous exploratory multiple imputation work using data from the ODYSSEY trial suggests that the number of cycles of the chained equations needs to be increased from the default of 10 to 30.

The choice of the number of imputations to be performed will be discussed in section 14.1.3.

14.1.1.2.1 SENSITIVITY ANALYSIS 1*

The multivariate imputation by chained equations procedure described in section 14.1.1.2 uses predictive mean matching to perform the imputation. Predictive mean matching uses normal linear regression to obtain linear predictions which are then used to match the missing values with the nearest neighbours. Since $\log_{10}$-transformed VL data are not likely to be normally distributed, the normality assumption of this model used by predictive mean matching might be violated. However, by drawing from the observed data to impute the missing data, predictive mean matching preserves the distribution of the observed VL measurements in the imputed data, which makes it more robust than using a linear regression imputation approach.

To explore sensitivity to violation of the normality assumption, a sensitivity analysis will be performed where, instead of imputing the $\log_{10}$-transformed VL measurement at each visit window, the indicator of suppression at each visit window (taking value 0 if HIV-1 RNA <50c/mL, 1 if HIV-1 RNA ≥50c/mL) is imputed. Left-censored VL measurements will be handled as per the procedure described in section 14.1.2, before the values of their associated indicators of suppression are derived.

One indicator of suppression per window will be included in the imputation model as a separate binary variable. For visit windows with more than 1 VL measurement, the choice of which measurement to include is described in section 14.1.1.3.

Imputation will be performed separately in each randomised arm (via mi impute chained, by() in Stata) to allow for any interactions with randomised arms.

Each missing indicator of suppression will be imputed using a logistic regression conditional imputation model (via mi impute chained (logit) in Stata), conditional on the following:

- observed and imputed indicators of suppression from all other visit windows;
- observed and mean imputed baseline $\log_{10}$-transformed VL data;
- an overall binary indicator of whether the participants made a change to their originally randomised treatment (as defined above);
- a variable containing the time from randomisation to treatment change (as defined above); and
- site.

This approach does not rely on the normality assumption, however it uses less information than modelling the distribution of $\log_{10}$-transformed VL data, and is prone to perfect prediction. In case of perfect prediction, an augmentation procedure will be performed where the dataset is augmented with a few extra observations to avert

perfect prediction, which are then assigned some small weights to limit their impact on the estimated imputation model[15]. This procedure has been implemented in the mi impute suite in Stata, via the option augment that is available with the logistic regression conditional imputation model. In case data augmentation cannot overcome perfect prediction in either arm (in more than 5% of the bootstrapped datasets, see section 14.1.3), the conditional imputation models will be simplified for both arms (and for all bootstrapped datasets, see section 14.1.3), so that the missing indicator of suppression in each visit window is imputed conditional on the following:

- the observed and imputed indicators of suppression from ± 1 adjacent window (via the option omit()in mi impute chained in Stata, e.g. indicator of suppression at week 48 will be imputed using indicators of suppression at weeks 40 and 60);
- observed and mean imputed baseline log_{10}-transformed VL data;
- an overall binary indicator of whether the participants made a change to their originally randomised treatment (as defined above);
- a variable containing the time from randomisation to treatment change (as defined above); and
- site.

14.1.1.3 Handling visit windows with more than 1 viral load measurement

For visit windows with more than 1 VL measurement, the measurement from the undiluted sample with date of collection closest to the scheduled visit date of that window (or if there are no undiluted samples, the diluted sample with closest date) will be selected for that window to be used in the imputation model. Where there are two values (meeting the same criteria) within a window, both equidistant from the nominal visit week, the later value will be used. Once missing VL measurements have been imputed for all scheduled visit windows, i.e. all windows have 1 measurement, the imputed data will be appended back to the originally observed data (for each of the B bootstrapped datasets, see section 14.1.3) in which some windows have multiple VL measurements. In this appended dataset, all log_{10}-transformed VL data will be back transformed to the original scale. The analysis of the primary endpoint will be performed on this appended dataset so that all available observed VL measurements as well as imputed VL measurements are used to derive the primary endpoint (the first of 2 consecutive HIV-1 RNA ≥50c/mL).

14.1.1.3.1 SENSITIVITY ANALYSIS 2*

A sensitivity analysis will be performed where the last undiluted log_{10}-transformed VL measurement of windows with multiple measurements will be selected for inclusion in the imputation model.

14.1.1.3.2 SENSITIVITY ANALYSIS 3*

Repeats following initial real-time HIV-1 RNA ≥50c/mL are mandated at all sites for weeks 48, and 96. A sensitivity analysis will be performed where, for each of weeks 48, and 96, 1 main log_{10}-transformed VL measurement and 1 repeat log_{10}-transformed VL measurement (if the initial VL is raised) will be selected, and missing repeats will be imputed as separate variables alongside VL measurements at all visit windows post-baseline as part of the *same* multiple imputation procedure described in sections 14.1.1.2 and 14.1.1.3 (i.e. 12 incomplete main VL variables and 2 incomplete repeat VL variables to be imputed together in one multivariable imputation by chained equations procedure).

Conditional imputation (via mi impute chained (pmm, cond()) in Stata) will be performed for the repeats, such that a missing repeat will only be imputed if the observed or imputed initial VL in the same visit window is raised. Each repeat variable will be imputed conditional on the following:

- the initial VL measurement in that window;
- the main measurement in the next window, e.g. missing repeat at week 48 will be imputed conditional on the initial VL in week 48 and VL in week 60 (via the option omit() in mi impute chained in Stata);
- observed and mean imputed log_{10}-transformed VL data at baseline;
- an overall binary indicator of whether the participants made a change to their originally randomised treatment (as defined above);
- a variable containing the time from randomisation to treatment change (as defined above); and

- site.

Repeats (i.e. the 2 repeat variables) are all excluded from the imputation model of missing VL measurements in all scheduled visit windows (i.e. the 2 repeats variables are not used to inform the imputation of the 12 main variables, via the option omit() in mi impute chained in Stata). Imputation will be performed separately in each randomised arm.

**Note: If 2 of the 3 sensitivity analyses described in sections 14.1.1.2.1, 14.1.1.3.1, 14.1.1.3.2 indicate sensitivity, they will be combined into a further sensitivity analysis.*

14.1.2. HANDLING VIRAL LOAD MEASUREMENT BELOW THE LIMIT OF DETECTION

In BREATHER Plus, samples are run on machines with highest LLOD = 50c/mL (e.g. LLOD = 20c/mL, 40c/mL, etc).

- If a sample is undiluted, the upper bound of the resulting VL measurement is equal to the LLOD (e.g. 20c/mL, 40 c/mL, 50c/mL);

- If a sample is diluted, the upper bound of the resulting VL measurement is equal to these values multiplied by the dilution ratio (1:2, 1:3, ..., 1:5), so the highest possible upper bound is 250c/mL.

A cut-off of <50c/mL is used to define suppression. Each sample either results in a VL reading ('x') or is known to be below the LLOD ('<x'), and the latter occurs with x = LLOD (e.g. <20c/mL, <40 c/mL, <50c/mL), or multiples of these limits when a dilution is made. This complicates both the primary analysis and the imputation of $\log_{10}$-transformed VL because for some observed VL measurements we only know the upper bound and not the actual values (i.e. VL data are left-censored). Suppression status (<50c/mL) is clear, except when the sample is known to be <LLOD *and* the LLOD is above 50c/mL due to dilution (see 3. below).

Below are some examples of potential scenarios given the 50c/mL cut-off for suppression:

Scenario 1 – Suppressed (<50c/mL):

a. Undiluted: observed HIV-1 RNA <20c/mL, <40c/mL, <50c/mL; VL measurement is left-censored;
b. Undiluted: observed HIV-1 RNA = 27c/mL, run on a machine with LLOD = 20c/mL; VL measurement is known;
c. Diluted: adjusted (for dilution) HIV-1 RNA <40c/mL, from a 1:2 diluted sample run on a machine with LLOD = 20 c/mL (so compared with adjusted upper bound = 40 c/mL); VL measurement is left-censored;
d. Diluted: adjusted HIV-1 RNA = 46c/mL, from 1:2 diluted sample run on a machine with LLOD = 20c/mL (so compared with adjusted upper bound = 40 c/mL); VL measurement is known;

Scenario 2 – Unsuppressed (≥50c/mL):

e. Undiluted: observed HIV-1 RNA = 105c/mL; VL measurement is known;
f. Diluted: e.g. adjusted HIV-1 RNA = 106c/mL, from a 1:2 diluted sample run on a machine with LLOD = 50c/mL (so compared with adjusted upper bound = 100c/mL); VL measurement is known;

Scenario 3 – Unclear:

g. Diluted: e.g. adjusted HIV-1 RNA <100c/mL, from a 1:2 diluted sample run on a machine with LLOD = 50c/mL (so compared with adjusted upper bound = 100c/mL); here we know that HIV-1 RNA is <100c/mL but we cannot determine if HIV-1 RNA is also <50c/mL or 50c/mL ≤ HIV-1 RNA < 100c/mL; VL measurement is left-censored.

The following procedure will be followed to deal with left-censored data.

Using all available undiluted samples at all visit windows considered in the analysis of the primary endpoint, we will calculate the percentage of participants who are virologically suppressed (HIV-1 RNA <50c/mL) among all participants whose HIV-1 RNA is known to be (i) <100/c/mL; and (ii) <200c/mL. The 200c/mL bound is chosen instead of the highest possible 250c/mL bound (for machines with LLOD = 50c/mL and samples diluted 1:5) since it is likely to occur more frequently in the dataset.

- If both percentages (i) and (ii) are ≥ 80%, all left-censored VL measurements (as described above) will be replaced with 25c/mL, based on the assumption that if the majority of participants whose VL measurements are <100c/mL and <200c/mL are in fact virologically suppressed then it is reasonable to replace with a measurement that reflects suppression;

- If percentage (i) is ≥ 80% while percentage (ii) is <80%, left censored VL measurements known to be either <50c/mL or <100c/mL will be replaced with 25c/mL; left censored VL measurements known to be <200c/mL (excluding those known to be <100c/mL) will be replaced with 100c/mL.

- If percentage (i) is <80% while percentage (ii) is ≥ 80% , or if both percentages (i) and (ii) are <80%, left-censored VL measurements known to be <50c/mL will be replaced with 25c/mL; left-censored VL measurements known to be <100c/mL (excluding those known to be <50c/mL) will be replaced with 50c/mL; and left-censored VL measurements known to be <200c/mL (excluding those known to be <100c/mL) will be replaced with 100c/mL.

14.1.3. COMBINING BOOTSTRAPPING WITH MULTIPLE IMPUTATION

Since bootstrapping will be used to estimate standard errors and confidence intervals in the primary analysis, multiple imputation needs to be performed in a way that allows for bootstrapping. We will use a procedure recommended by Bartlett & Hughes (2020) involving bootstrapping followed by imputation (termed Boot-MI)[16], where:

- B bootstrapped datasets are created from bootstrapping the incomplete dataset. Bootstrapped samples are selected within strata defined by randomised arms and stratification variables;

- Each bootstrapped dataset is imputed M times using the procedure described in sections 14.1.1 and 14.1.2, resulting in B×M bootstrapped imputed datasets;

- The substantive analysis is performed in each of these B×M datasets, producing B×M point estimates (adjusted difference in proportions);

- The pooled point estimate for inference is the average of these B×M estimates; a one-way random effects ANOVA model is fitted to these B×M point estimates and its results used to estimate the associated standard error; 95% confidence interval is based on a *t*-distribution.**

***Note: it is not possible to obtain bias-corrected bootstrap confidence intervals with Boot-MI, since the confidence interval is based on a t-distribution.*

Bartlett & Hughes (2020) showed that using a large number of bootstrapped datasets (e.g. B=1000) and a small number of imputations (M=2) is computationally efficient[16]. For BREATHER Plus, B=10,000 bootstrapped datasets and M=5 imputations will be created, resulting in 50,000 bootstrapped imputed datasets that will be analysed according to the Boot-MI procedure described above.

14.1.4. USING IMPUTED MISSING VIRAL LOAD MEASUREMENTS AFTER LOSS TO FOLLOW-UP AND DEATH

Loss to follow-up and death are not components of the primary endpoint in BREATHER Plus (proportion of children with confirmed viral rebound, defined as the first two consecutive HIV-1 RNA ≥50c/mL, by week 96).

Multiple imputation will be performed for all visit windows as described above. Then for participants who were lost to follow-up, withdrew, or died prior to meeting the primary endpoint, if their last observed or imputed VL measurement is raised (HIV-1 RNA ≥50c/mL) prior to loss to follow-up, withdrawal, or death, this measurement as well as the VL measurement imputed in the next visit window will be used to derive the primary endpoint or to determine the censoring time. All subsequent imputed VL measurements for these participants are not used for this purpose and will be deleted. The procedure is summarised in Appendix Table A1.

The imputation procedure described in section 14.1.1 does not include as covariates indicators of loss to follow-up, withdrawal, and death. Since none of these events is a component of the primary endpoint (proportion of children with confirmed viral rebound, defined as the first two consecutive HIV-1 RNA ≥50c/mL, by week 96), omitting these variables from the imputation model will likely not bias the results.

Appendix Table A1. Procedure for using VL measurement observed before loss to follow-up/death and imputed VL measurement at the following visit window to derive the primary endpoint or determine the censoring time.

Observed or imputed VL at week *w* visit window		**Imputed VL at week *w+12*[ǂ] visit window**	**Status**
≥50c/mL (with no confirmation)	The participant was lost to follow-up, withdrew, or died between week *w* and *w+12*[ǂ]	≥50c/mL	The participant met the endpoint at week *w*
≥50c/mL (with no confirmation)		<50c/mL	The participant did not meet the endpoint by week *w+12*[ǂ], censored at week *w+12*[ǂ]
<50c/mL		≥50c/mL	The participant did not meet the endpoint by week *w*, censored at week *w*
<50c/mL		<50c/mL	The participant did not meet the endpoint by week *w*, censored at week *w*

[ǂ]or *w+8* if *w≤40*.

Note that the methods described in Appendix 1 were written (initially for the D3 trial) by Dr Tra My Pham with support from Professor Ian White without reference to accumulating unblinded data. Dr Man Chan and Dr Debbie Ford (who are unblinded to accumulating data) reviewed the methods but did not provide information on D3 or BREATHER Plus data.

14.2 APPENDIX 2 - PROPOSED RULES FOR REQUESTING RETROSPECTIVE HIV-1 RNA TESTING FROM STORED PLASMA SAMPLES

To avoid requesting retrospective samples for a given visit, which already have real time HIV-1 RNA result available, the following rules will be used:

Appendix table A2. Proposed rules for requesting retrospective HIV-1 RNA testing from stored plasma samples

Real time HIV-1 RNA	**Retrospective HIV-1 RNA** (Stored plasma sample)
Available In database for a visit week window	**DO NOT REQUEST stored sample testing** (unless dilution has been used and result cannot be classified as < or >=50 c/mL) [in visit week window]
Not available In database for a visit week window	
Pending data entry Checked and confirmed by site real time HIV-1 RNA sample taken at visit; pending entry	**DO NOT REQUEST stored sample testing** [in visit week window]
Confirmed missing HIV-1 RNA Checked and confirmed by site real time HIV-1 RNA sample is not expected (e.g. due to missed visit, sample lost, failed test ,etc)	**REQUEST stored sample testing** [in visit week window]

Regularly extract HIV-1 RNA data from database and keep track of availability of real time HIV-1 RNA data.

14.3 APPENDIX 3 - CONSIDERATIONS/ASSUMPTIONS ON HIV-1 RNA DATA FOR ANALYSIS

Real time and retrospective HIV-1 RNA

Real time HIV-1 RNA will be used for analysis where available at each study visit (mandatory at weeks 48, 96 and 48 weekly thereafter and end of trial visit). When a real time result is not available (including if missing at a mandatory visit or confirmatory timepoint), a retrospective HIV-1 RNA will be requested and used for the analysis. Similarly, if for a real-time result, the HIV-1 RNA Lower Limit of Detection (LLD) is >50 c/mL (for example, when a HIV-1 RNA test is conducted on a diluted sample or dried blood spot) a retrospective HIV-1 RNA test will be requested.

Where a participant has a real time and retrospective HIV-1 RNA within the same visit week window, the real time HIV-1 RNA will be used as the retrospective result should not have been requested. The exception to this is when the real time HIV-1 RNA Lower Limit of Detection (LLD) is >50 c/mL (as above), in which case, the retrospective sample will be used instead.

Diluted samples

Sites/labs must clearly indicate the dilution factor used if samples were diluted for HIV-1 RNA measurement.

HIV-1 RNA measured in diluted samples may be provided in two forms depending on site:

1) **Diluted HIV-1 RNA result (unadjusted for dilution factor; dilution factor not automatically adjusted for by the assay instrument):** these results will be adjusted for dilution factor when creating analysis datasets by multiplying the result by the dilution factor to obtain the "useable" result i.e. result adjusted for dilution. This "useable" result will be used in the analysis.

2) **Undiluted HIV-1 RNA result (adjusted for dilution factor; dilution factor already automatically adjusted for by the assay instrument**): these results are already in "useable" form and will be used in the analysis as provided.

For the analysis of primary endpoint (confirmed HIV-1 RNA ≥50 c/mL), the following considerations will be implemented for grouping of HIV-1 RNA results into <50 c/mL and ≥50 c/mL groups using the "useable" HIV-1 RNA results:

1) **"useable" HIV-1 RNA result with value:**
   - **a.** equal to or greater than ("≥") 50 c/mL and sign "=" will be allocated to "≥50 c/mL" group (e.g. HIV-1 RNA=80 participant will be allocated to ≥50 c/mL);
   - **b.** less than ("<") 50 c/mL and sign "=" or "<" will be allocated to "<50 c/mL" group;
   - **c.** equal to ("=") 50 c/mL and sign "<" will be allocated to "<50 c/mL" group;

2) "useable" HIV-1 RNA result with value greater than (">") 50 c/mL and sign "<", for example, <80, <100 and <200, will not be used for grouping and will be assumed to be missing.

3) If "useable" HIV-1 RNA result is missing due to target not detected, it will be replaced by the LLD with a sign "<". For example, the LLD=40 c/mL, HIV-1 RNA = "target not detected" and sign is N/A, we will replace "target not detected" by 40 and assign sign "<". Once replaced, grouping will be done as described above in points 1) and 2).

14.4 APPENDIX 4 - DETAILS OF MODIFIED FDA SNAPSHOT ALGORITHM

This section outlines the modified version of the FDA snapshot algorithm, which will be used to compare virological rebound (HIV-1 RNA ≥50 c/mL) in SCT vs. CT at weeks 48 and 96.

Since the BREATHER Plus protocol specifies the following drugs are interchangeable, switches from 3TC to FTC and TDF to TAF (or vice versa) will be ignored in the algorithm. Switch for lack of efficacy will always be treated as a non-permitted change, however, when a participant switches from 3TC to FTC and TDF to TAF (or vice versa) for this reason, they will likely switch another component of their regimen at the same time.

Non-permitted changes in SCT arm include:

- Switch from SCT to continuous therapy for any reason (excluding switch for protocol deviation or patient/carer decision <7 days [permitted change, see below]).

Non-permitted changes in both arms include:

- Change of ART component(s) due to lack of efficacy
- Change of ART component(s) due to adverse event*
- Change of ART component(s) due to pregnancy or desire to become pregnant*
- Change of ART component(s) due to protocol deviation or patient/carer decision where the time off the allocated regimen is ≥7 days*

*Except for changes from 3TC to FTC and TDF to TAF (or vice versa) which are ignored

Permitted ART changes in both arms include:

- Change of ART component(s) or strategy for incorrect prescribing (protocol deviation) or patient/carer decision provided that the participant switches back to their allocated regimen and strategy <7 days after change*

*Except for changes from 3TC to FTC and TDF to TAF (or vice versa) which are ignored

Ignored ART changes in both arms include changes where components of the ART regimen remain unchanged (and in the SCT arm, the strategy remains unchanged) unless otherwise specified:

- Dose change of 3TC for creatinine clearance
- Move to double DTG dose (or adjustment to TAF dose) for TB treatment and return to single dose following completion of TB treatment or similar dose adjustments due to drug-drug interactions with other concomitant medications (applicable in CT arm only)
- Switch to fixed dose ART combination (FDC) (or FDC to single/dual) containing the same ART components
- Changes between different products containing the same ART components
- Changes between mornings and evenings
- Switch from 3TC to FTC, or vice versa, for any reason other than lack of efficacy
- Switch from TDF to TAF, or vice versa, for any reason other than lack of efficacy

Classification (ignored/permitted/non-permitted) of ART changes for any other reason will be reviewed by a clinician blinded to trial arm and with no access to viral loads. ART switches for breastfeeding (not directly captured on ART log but via “Other” code) will be classified according to rules used for pregnancy.

ART interruptions/stops

Participants who have not been seen since baseline and have no HIV-1 RNA data on ART post-baseline will not be categorised for the purposes of the FDA snapshot algorithm and will therefore be excluded from the analysis.

Only post-baseline and HIV-1 RNA on-ART will be used in the algorithm. Stop of any ART component is considered an ART interruption. Participants interrupting ART prior to or within week 48/96 window are not classified as treatment switch until a new treatment has started (rules above will be applied). If they resume the same treatment, the period of the ART interruption is ignored although HIV-1 RNA off ART will not be used.

Where a participant has <7 days on an incorrect/new regimen/strategy for protocol deviation or patient/carer decision before returning to their allocated regimen/strategy, HIV-1 RNA data will not be used in the algorithm until the participant has been back on their allocated regimen and strategy for ≥7 days. HIV-1 RNA data post change are not used in the algorithm for participants who have ≥7 days on an incorrect/new regimen/strategy for protocol deviation or patient/carer decision.

Multiple HIV-RNA results within a window

In the case where a participant has multiple HIV-1 RNA results within week 48/96 window, the latest HIV-1 RNA should be used in the FDA snapshot algorithm.

Secondary analysis

Given the trial protocol stipulates SCT participants must switch to continuous therapy when becoming pregnant, receiving treatment for TB (or other dose adjustments due to drug-drug interactions with other concomitant medications), a secondary analysis will treat switches from SCT to CT due to pregnancy/breastfeeding or TB treatment/other concomitant medications as a permitted switch.

Appendix table A3. Full algorithm for modified FDA snapshot analysis of virological rebound

1. Non-permitted change in therapy prior to week 48/96		
1a. Last on-treatment HIV-1 RNA at/prior to change ≥50 c/mL (or change for lack of efficacy*)	HIV-1 RNA≥50 c/mL	Change in therapy/strategy while HIV-1 RNA ≥50 c/mL
1b. Last on-treatment HIV-1 RNA at/prior to change <50 c/mL or no on-treatment HIV-1 RNA available during study	No virologic data in week 48/96 window	Change in therapy/strategy while HIV-1 RNA <50 c/mL
2. Permitted change in therapy prior to week 48/96 and last HIV-1 RNA on treatment at/prior to change ≥50 c/mL	HIV-1 RNA≥50 c/mL	Change in therapy/strategy while HIV-1 RNA ≥50 c/mL
3. Non-permitted change in therapy during week 48/96		
3a. Last on-treatment HIV-1 RNA during week 48/96 prior to/on the date of change ≥50 c/mL	HIV-1 RNA≥50 c/mL	Data in window not below 50 c/mL
3b. Last on-treatment HIV-1 RNA during week 48/96 prior to/on the date of change <50 c/mL	HIV-1 RNA<50 c/mL	Data in window below 50 c/mL
3c. No HIV-1 RNA during week 48/96 prior to/on the date of change		
3c (i). Last on-treatment HIV-1 RNA at/prior to change ≥50 c/mL (or change for lack of efficacy)	HIV-1 RNA≥50 c/mL	Change in therapy/strategy while HIV-1 RNA≥50 c/mL
3c (ii). Last on-treatment HIV-1 RNA at/prior to change <50 c/mL or no on-treatment HIV-1 RNA available during study	No virologic data in week 48/96 window	Change in therapy/strategy while HIV-1 RNA<50 c/mL
4. Permitted change in therapy during week 48/96 and last on treatment HIV-1 RNA at/prior to change ≥50 c/mL		
4a. Last on-treatment HIV-1 RNA is during week 48/96 and prior to/on date of change	HIV-1 RNA≥50 c/mL	Data in window not below 50 c/mL (see 3a)
4b. Last on-treatment HIV-1 RNA is prior to week 48/96	HIV-1 RNA≥50 c/mL	Change in therapy/strategy while HIV-1 RNA≥50 c/mL (see 3c)
If none of above		

5. HIV-1 RNA available on allocated regimen/strategy (with prior permitted changes allowed while HIV-1 RNA<50 c/mL) in 48/96 week window		
5a. Last on-treatment HIV-1 RNA during week 48/96 ≥50 c/mL (assuming anyone with HIV-1 RNA≥50 c/mL at week 48/96 will be tested by week 54/102**)	HIV-1 RNA≥50 c/mL	Data in window not below 50 c/mL
5b. Last on-treatment HIV-1 RNA during week 48/96 <50 c/mL (assuming anyone with HIV-1 RNA≥50 c/mL at week 48/96 will be tested by week 54/102**)	HIV-1 RNA<50 c/mL	Data in window below 50 c/mL
6. Participant is on allocated regimen/strategy (including permitted changes while HIV1-RNA<50 c/mL) but has no HIV-1 RNA data in 48/96 week window	No virologic data in week 48/96 window	Missing data in window
7. Participant died/lost to follow-up/withdrew prior to 48/96 week window and last on-treatment viral load <50 c/mL or no on-treatment HIV-1 RNA available during study	No virologic data in week 48/96 window	Death, LTFU, withdrawal from trial while HIV1-RNA<50 c/mL
8. Participant died/lost to follow-up/withdrew prior to 48/96 week window and last on-treatment viral load ≥50 c/mL	HIV-1 RNA≥50 c/mL	Death, LTFU, withdrawal from trial while HIV-1 RNA≥50 c/mL

***** Sites will be asked to do a viral load test before changing treatment for suspected treatment failure; however, if there is a treatment change for clinical or immunological failure, this will be included here irrespective of most recent viral load

******Every effort will be made to do confirmatory viral loads within the visit window. If this is not done for any reason, then the participant will be classified under 5a and not 5b (i.e. as HIV-1 RNA≥50 c/mL).

Ignored changes do not affect participant's classification.

Any viral loads which due to dilution cannot be classified as </≥50 c/mL will not be used.

Appendix table A4. Categorisation of Virological Outcomes at 48/96 weeks (within window 42-54/90-102 weeks)

Ignored, permitted, and non-permitted ART changes are specified above. If participant has changed regimen/strategy in any way in either arm (including loss to follow-up/withdrawal or death) before/during week 48/96 then **(i)** if last HIV-1 RNA prior to change is ≥50 c/mL, they are a rebound, **(ii)** if last HIV-1 RNA <50 c/mL they will have no virologic data unless they made a permitted change in which case 48/96 week data can be used as if they are continuing as randomised.

	SCT	**CT**
	N (%)	N (%)
HIV-1 RNA≥50 c/mL[1]		
Treatment difference (95% CI)		
HIV-1 RNA<50 c/mL[2]		
No virological data in week 48/96 window		
Discontinued allocated regimen/strategy due to AE or death and last on treatment (at/prior to change) HIV-1 RNA<50 c/mL[3]		
Discontinued allocated regimen/strategy for other reasons and last on treatment (at/prior to change) HIV-1 RNA<50 c/mL[4]		
On study allocated regimen and strategy but missing HIV-1 RNA data in window[5]		

[1] Includes: **(i)** participants on allocated regimen and strategy (SCT/CT), including those with prior permitted changes while HIV-1 RNA<50 c/mL, who had confirmed HIV-1 RNA≥50 c/mL in 48/96 week window; **(ii)** participants who changed any component of allocated regimen or resumed continuous therapy in the SCT arm because of lack of efficacy prior to/during week 48/96 window; **(iii)** participants who discontinued/changed any component of allocated regimen or resumed continuous therapy in the SCT arm for reasons other than lack of efficacy prior to/during week 48/96 with the last on treatment (prior to/on the date of change) HIV-1 RNA ≥50 c/mL.

[2] Includes: **(i)** participants on allocated regimen and strategy (SCT/CT), including those with prior permitted changes while HIV-RNA<50 c/mL, with HIV-1 RNA<50 c/mL in week 48/96 window.

[3] Includes participants who discontinued or changed any component of allocated regimen or strategy for AE or death before or during week 48/96 where last on treatment HIV-RNA<50 c/mL.

[4] Includes participants who discontinued or changed any component of allocated regimen or strategy for reasons other than an AE or death or lack of efficacy, e.g., withdrew consent, lost to follow-up, pregnancy (or desire to become pregnant), transferred care to a non-study site, patient/carer decision or protocol deviation that do not meet the criteria for permitted change, or had any other non-permitted change before or during week 48/96 where last on treatment HIV-1 RNA<50 c/mL.

[5]Includes participants remaining on allocated regimen and strategy (SCT/CT), including those with prior permitted changes while HIV-1 RNA<50 c/mL, who had no available HIV-1 RNA in week 48/96 window.

Ignored changes do not affect participant's classification.
Any viral loads which due to dilution cannot be classified as </≥50 c/mL will not be used.

REFERENCES

1. Quartagno, M., et al., *Handling an uncertain control group event risk in non-inferiority trials: non-inferiority frontiers and the power-stabilising transformation.* Trials, 2020. **21**(1): p. 145.
2. Matilde Sanchez, M. and X. Chen, *Choosing the analysis population in non-inferiority studies: per protocol or intent-to-treat.* Stat Med, 2006. **25**(7): p. 1169-81.
3. Sheng, D. and M.Y. Kim, *The effects of non-compliance on intent-to-treat analysis of equivalence trials.* Stat Med, 2006. **25**(7): p. 1183-99.
4. Wiens, B.L. and W. Zhao, *The role of intention to treat in analysis of noninferiority studies.* Clin Trials, 2007. **4**(3): p. 286-91.
5. World Health Organisation, *WHO case definitions of HIV for surveillance and revised clinical staging and immunological classification of HIV-related disease in adults and children (http://www.who.int/hiv/pub/guidelines/HIVstaging150307.pdf)*. 2007.
6. *U.S. Department of Health and Human Services, National Institutes of Health, National Institute of Allergy and Infectious Diseases, Division of AIDS. Division of AIDS (DAIDS) Table for Grading the Severity of Adult and Pediatric Adverse Events, Corrected Version 2.1. [July 2017]. Available from: https://rsc.niaid.nih.gov/sites/default/files/daidsgradingcorrectedv21.pdf.*
7. WHO, *Antiretroviral therapy for HIV infection in infants and children: Towards universal access. Recommendations for a public health approach: 2010 revision* Vol. 13 July 2010. 2010, Geneva.
8. Morris, T.P., et al., *Planning a method for covariate adjustment in individually-randomised trials: a practical guide. arXiv:2107.06398.* https://arxiv.org/, 2021.
9. Royston, P. and M.K. Parmar, *Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects.* Statistics in medicine, 2002. **21**(15): p. 2175-2197.
10. *Food Drug Administration Center for Drugs Evaluation Research. Human Immunodeficiency Virus-1 Infection: Developing Antiretroviral Drugs for Treatment - Guidance for Industry.* 2015: Maryland, USA: FDA.
11. Mary E. Nilsson, et al., *https://cssrs.columbia.edu/wp-content/uploads/ScoringandDataAnalysisGuide-for-Clinical-Trials-1.pdf*. 2013.
12. White, I.R. and S.G. Thompson, *Adjusting for partially missing baseline measurements in randomized trials.* Statistics in medicine, 2005. **24**(7): p. 993-1007.
13. Van Buuren, S., H.C. Boshuizen, and D.L. Knook, *Multiple imputation of missing blood pressure covariates in survival analysis.* Statistics in medicine, 1999. **18**(6): p. 681-694.
14. Rubin, D.B., *Multiple imputation for nonresponse in surveys Donald B. Rubin*. Wiley classics library. 2004, Hoboken, N.J. ;: Wiley-Interscience.
15. White, I.R., R. Daniel, and P. Royston, *Avoiding bias due to perfect prediction in multiple imputation of incomplete categorical variables.* Computational statistics & data analysis, 2010. **54**(10): p. 2267-2275.
16. Bartlett, J.W. and R.A. Hughes, *Bootstrap inference for multiple imputation under uncongeniality and misspecification.* Statistical methods in medical research, 2020. **29**(12): p. 3533-3546.

BREATHER PLUS STATISTICAL ANALYSIS PLAN APPENDIX - SAP UPDATES FOR FINAL ANALYSIS

SAP Appendix version and approvals

Version Number and Date: 1.0 (19 May 2025) Supersedes version: 0.8 (19 May 2025) **In relation to SAP version:** 2.0			
Author	**Position**	**Signature**	**Date**
Angus Jennings	Delegated Statistician, MRC CTU at UCL		
Reviewed by			
Dr Deborah Ford	Trial Statistician (Senior Statistician), MRC CTU at UCL		
Professor Rodolphe Thiébaut	Independent Senior Statistician (reviewer), University of Bordeaux		
Professor Ian White	Independent Senior Statistician (reviewer), MRC CTU at UCL		

CONTENTS

This document describes decisions made after SAP v2.0 was signed off and the final 96-week data were available. It was reviewed by methodologists Ian White and Matteo Quartagno, MRC CTU at UCL, who were blind to study data. Angus Jennings and Deborah Ford had seen 96-week data.

1. Primary analysis testing scheme

The BREATHER plus SAP V2.0, section 10.13, states:

> *The primary analysis is non-inferiority of SCT versus CT. If the confidence interval for the treatment effect (difference in proportion experiencing confirmed viral rebound by 96 weeks (SCT-CT)) lies below the selected non-inferiority margin (selected according to the SAFE non-inferiority frontier described in Section 6.1), then we will also test for superiority; because this is a closed test procedure there is no issue of multiplicity.*

The significance level for superiority testing was not stated. The plan for subsequent testing of inferiority in the case where the null hypothesis was not rejected (and hence non-inferiority ***not*** accepted) was also not described.

Dr Matteo Quartagno and Prof. Ian White were consulted (blind to trial data) regarding the appropriate significance level to use for testing of superiority or inferiority. Both agreed that the typical 5% significance level was appropriate irrespective of the significance level used to test for non-inferiority. The use of the 1% significance level for a control event rate below 9% when using the SAFE NI margin is only required to conserve the nominal Type I error rate at ~5% because the NI margin used is data-driven[1], and the subsequent superiority or inferiority test is independent of the variable NI margin.

For a closed testing procedure under rejection of the null hypothesis, the testing of each hypothesis should be individually controlled to 5%; thus, superiority of SCT to CT can be considered at the 5% significance level (using a 95% CI), where inferiority by a pre-specified margin has been rejected. While the closed test procedure does not apply to a test for inferiority, where inferiority by a pre-specified margin has not been rejected, it was agreed that a conservative approach (i.e. 5% significance level) was also appropriate for testing for harm (based on a two-sided test of the null hypothesis that the difference between arms (SCT-CT) =0).

2. NI margin for non-inferiority testing of analyses other than the primary analysis (of the primary endpoint)

The NI margin to use for primary analysis of the primary endpoint is fully pre-specified in the SAP, in line with the previously published SAFE frontier[1]. The SAP also specifies a number of sensitivity analyses (analyses unadjusted for site, accounting for non-proportional hazards, or using unimputed data if multiple imputation is used for the primary analysis) and supplementary analyses of the primary endpoint (per-protocol analysis and a composite version of the primary endpoint) as well as a secondary endpoint considering VL≥1000 (see list in Table 1). Several of these are referred to as being analysed ‘as described for the primary analysis’, without describing the significance level or an NI margin, if appropriate.

As the NI margin was only pre-specified for the difference between arms reaching the primary endpoint (confirmed VL≥50 c/mL by week 96), we consider that the NI frontier agreed with clinical colleagues should not be extrapolated to other clinical endpoints. Thus, (1) the composite endpoint considering the first of the primary endpoint or any switch from SCT to CT

other than for pregnancy as an event; and (2) the secondary endpoint of confirmed VL≥1000 c/mL by 96 weeks are excluded from NI testing and a 5% significance level (for testing of superiority/inferiority) will be considered in line with SAP wording (section 10.1), *'[s]tatistical tests will be 2-sided, estimates will be presented with a 2-sided 95% CI'*. In addition, no NI margin was defined for hazard-based comparisons (the secondary Cox analysis of the primary endpoint and the sub-group analysis of the primary endpoint, see section 5), and hence the SAFE frontier presented in the SAP was never intended for use with these analyses.

For alternate analyses of the difference in proportion of participants meeting the primary endpoint, namely the three sensitivity analyses and the per-protocol analysis, the NI frontier is still considered appropriate; for sensitivity analyses, only the analysis method/assumptions have been changed (and the estimand is unchanged), and for the per-protocol analysis only the population/intercurrent event strategy is changed. For these analyses, the full analysis will be re-considered following the methodology used for the primary analysis, i.e. the NI margin will be recalculated based on the new control group event rate and the appropriate significance level will be chosen based on this.

3. Confidence intervals for estimates of the proportion with confirmed VL≥50/1000 by trial arm

The SAP, section 10.8, incorrectly implies use of a 99% CI for estimates of the proportion of participants with viral rebound in each treatment group, SCT/CT.

> *The 95% CI or 99% of the estimated probabilities of confirmed viral rebound by week 96 in each randomised group will also bias-corrected.*

This was a typographical error when the SAP was written. Since there is no data-dependent threshold for comparison (equivalent to the NI margin determined by control group event rate), there is no rationale for adjusting the significance level to maintain the Type I error rate; a 95% confidence interval will be used.

Table 1. ***Agreed analysis methods for analyses of proportions with confirmed VL≥50/1000***

Efficacy endpoint[1]	**Analysis[1]**	**2-sided significance level *α*, *(1-α)*100% CIs***		
		Estimates by trial arm[2]	Treatment effect estimates	
			Non-inferiority framework	*Superiority/inferiority framework*[3]
Primary endpoint [confirmed VL≥50]	Primary analysis	5%	5%, or 1% if CT rebound rate < 9%	5%
	Sensitivity: *unadjusted analysis*	5%	5%, or 1% if CT rebound rate < 9%[4]	5%
	Sensitivity: *non-PH analysis*	5%	5%, or 1% if CT rebound rate < 9%[4]	5%
	Sensitivity: *non-imputed data [if MI required]*	5%	5%, or 1% if CT rebound rate < 9%[4]	5%
	PP analysis	5%	5%, or 1% if CT rebound rate < 9%[4]	5%
Analysis of composite definition of primary endpoint [including switch for other reasons]		5%	NA [no NI comparison to be made][5]	5%
Analysis of secondary endpoint [confirmed VL≥1000]		5%	NA [no NI comparison to be made][5]	5%

[1]Endpoints/analyses included where deemed to be lack of clarity in SAP. Significance level is specified in the SAP for hypothesis testing based on the Cox model (secondary analysis of primary endpoint), cross-sectional VLs and FDA snapshot.
[2]Whilst the SAP implies use of a 99% CI when CT risk<9% (i.e. two-sided significance level 1%) for estimates by trial arm (for at least the primary analysis of the primary endpoint but, by extension, other analyses completed 'as described for' the primary analysis), it was agreed that this was stipulated in error and that 95% CIs are more appropriate, see above (section 3).
[3]Significance level for inferiority/superiority not previously specified in SAP, refer to section above (section 2).
[4]When using the same endpoint/summary measure as the primary analysis of the primary endpoint (i.e. difference in proportion with confirmed VL≥50 at 96 weeks), it was agreed that the significance level and associated non-inferiority margin would be adjusted per the SAFE framework in the same way as for the primary analysis of the primary endpoint, but based on the observed control event risk in the respective analyses; see section above (section 2).
[5]The SAP stipulates analysis as described for the primary analysis, implying (but not explicitly stating) that significance level would depend on observed CT failure rates, however it was agreed that for any endpoint/summary measure other than the primary endpoint, the NI frontier was not pre-specified and hence a non-inferiority comparison was not appropriate.

4. Excluded switches for the composite endpoint (supplementary primary analysis)

The BP SAP V2.0, section 10.8, states:

> *In addition, since there may be other reasons for return from SCT to CT (e.g. persistent viral blips or clinician concern about adherence), we will consider a failure endpoint comprising HIV-1 RNA ≥50 c/mL (confirmed on a separate sample) or switch off randomised treatment regimen or strategy (except for treatment/strategy change for pregnancy) at any time up to end of week 96 window.*

It was decided that any reasons for switch described in free text that directly correspond to 'Wants to conceive/not using highly effective contraception' would also be excluded from the composite endpoint since, in the context of this endpoint, which aims to capture *all* cases of treatment/strategy failure, these switches are deemed to be of a similar nature to switch due to pregnancy and hence not of relevance.

More generally, switches from SCT to CT for reasons of 'Close-out visit' or free text that directly corresponds to 'Study withdrawal' (provided withdrawal is unrelated to strategy) are excluded from all summaries/analyses based on switches from SCT to CT, as the reporting of switches for these reasons is administrative.

5. Analysis plan for sub-group analyses of primary endpoint

Stratified treatment effects used in forest plots are calculated from pooled Cox models, with a main effect and a treatment-interaction effect included for each subgroup variable in its respective model (i.e. site or baseline age group). Considering heterogeneity by age group, unadjusted and adjusted models will be considered; in the adjusted model, main effects are also included for site. The interaction p-value presented is from the test of equality to 0 of all interaction terms.

An alternate, exploratory, sub-group analysis will be considered that is more comparable to the primary endpoint, using flexible parametric models to estimate the proportion with confirmed viral rebound in each sub-group, as described for the non-PH sensitivity analysis (provided model convergence is achieved given small event counts in sub-groups).

95% CIs will be used in all cases.

6. Visit windows used for assessments required at close-out

The BP SAP V2.0, section 9.3, states:

> *Laboratory measurements, and other clinical parameters (e.g. weight), at any nominal week are defined as those taken nearest to the nominal week within equally spaced windows according to the protocol visit schedule (note that the visit schedule changes frequency). The midpoint between two scheduled visit weeks should be taken as belonging to the latter window (see below). Where there are two values within one of these equally spaced windows, but both equidistant from the nominal visit week, the later value will be used.*
>
> *[…]*

For laboratory measurements or assessments taken less frequently in line with the protocol schedule, the nominal week windows will be widened according to the above principle, as determined by scheduled visits and midpoint between scheduled visits. For measurements taken every 48 weeks, a rule of +/- 24 weeks defining intervals will be applied; measurements taken every 24 weeks, a rule of +/- 12 weeks will be applied; while measurements taken every 16 weeks, a rule of +/- 8 weeks will be applied.

There are a number of measurements that are not completed at every ***scheduled*** visit beyond week 96 but will be completed at the ***close-out visit*** (namely vital signs, C-SSRS, mood survey, HATQoL, EQ-5D). It was decided that in the case of a close-out visit, +/-6-week windows would be used for summary and analysis, to correctly assign the close-out measurements to their respective scheduled week, and avoid assessments completed between week 96 and week 144 (and not within 6 weeks of either) being assigned as week 96/144 assessments.

7. Selection of lab measurements when multiple in an analysis window

When selecting the lab measurement to include in summaries/analyses by scheduled visit when there are multiple results within a visit window, the selection is done at the eCRF level, rather than the metabolite level. That is, if, of all the T-cell analyses conducted on a sample taken on a participant's week 96 target date, only the lymphocyte result was not available, only the available results from the week 96 date would be used in analysis (i.e. lymphocyte taken as missing), rather than using an alternate result that may be available in the window. This is to ensure relationships between metabolites within eCRFs are conserved. The full list of metabolites being considered together is summarised[1].

REFERENCES

1. Quartagno, M., et al., *The Smooth Away From Expected (SAFE) non-inferiority frontier: theory and implementation with an application to the D3 trial.* Trials, 2023. **24**(1): p. 556.

[1] **eCRF** (metabolites); **haematology** (haemoglobin, white blood cell count, absolute lymphocytes, absolute neutrophils, platelets), **biochemistry** (creatinine [& eGFR], albumin, ALT, total bilirubin), **lipids** (total cholesterol, HDL cholesterol, LDL cholesterol, triglycerides, phosphate), **T-cell lymphocyte subsets** (CD4 %, absolute CD4, CD8 %, absolute CD8, total lymphocytes).