**Campath, Calcineurin inhibitor reduction and Chronic allograft nephropathy**

# Data Analysis Plan for assessing clinical efficacy and safety of sirolimus-based maintenance treatment in the 3C Study

## 1   Background

This Data Analysis Plan describes the strategy, rationale and statistical methods that will guide assessment of the clinical efficacy and safety of sirolimus-based maintenance therapy in the 3C Study. A separate Data Analysis Plan will describe the assessment of Campath-based induction therapy. All analyses and reports will be prepared by the coordinating centre in the Clinical Trial Service Unit, University of Oxford.

The 3C Study is a randomized trial investigating two strategies to improve long-term outcomes in kidney transplantation. Firstly, it is comparing Campath (alemtuzumab) with basiliximab as the basis of induction therapy. Secondly, it is comparing an elective conversion to sirolimus-based maintenance therapy at around 6 months after transplantation compared to remaining on tacrolimus-based maintenance therapy. Follow-up visits are scheduled at 1, 3, 6, 9 and 12 months after transplantation and will then be followed by annual mailed questionnaires. In addition, all participants will be flagged with various NHS registries.

## 2   Comparisons of sirolimus-based versus tacrolimus-based maintenance therapy

All comparisons will involve comparing outcome during the defined analysis period among all those participants allocated at the maintenance randomization to receive sirolimus-based induction versus all those allocated to receive tacrolimus-based maintenance therapy (i.e. "intention-to-treat" analyses).[1, 2] All analyses will be stratified by induction therapy arm.

### 2.1  Primary comparison

The primary comparison will be of graft function (estimated using the 4 variable MDRD formula) at 2 years after transplantation. Any surviving participants with no valid values of creatinine (and hence eGFR) in this time window will be defined as having missing values and imputed as described in section 3.1 below). Values at 3, 6, 9, 12 and 15 months after randomization will also be presented but not compared.

### 2.2  Secondary comparisons

The secondary comparisons will be of the incidence of the first occurrence (after maintenance randomisation) of:

   i.   Graft outcomes:
        a.   Graft rejection

      i.   Biopsy-proven acute rejection: cellular (subdivided into Banff 1, 2, 3 and unknown), humoral (subdivided into Banff 1, 2, 3 and unknown) and all[3]

      ii.   All rejection: presumed and biopsy-proven (as above)

  b.  Graft survival (with reasons for failure subdivided into glomerular disease, fibrosis/atrophy, medical/surgical condition, rejection or unknown)

ii.   Safety outcomes:

  a.  Serious infections

      i.   All serious infections: subdivided into opportunistic and non-opportunistic

      ii.   Opportunistic infections subdivided into CMV, BK, fungal (non-invasive and invasive) and other (PCP, mycobacterial and other)

      iii.   Non-opportunistic subdivided into urinary tract, respiratory, gastrointestinal, central nervous system, other/unknown

  b.  Cancer

      i.   All cancer: subdivided into haematological (including post-transplant lymphoproliferative disorder and other); skin (subdivided into non-melanoma and melanoma); lung; gastrointestinal; hepatobiliary; breast; prostate; other/unspecified

iii.   Other outcomes of interest

  a.  Major vascular events (composite of non-fatal myocardial infarction, cardiac death, non-fatal or fatal stroke and arterial revascularisation)

  b.  Death: subdivided into vascular, infection, cancer and other causes

  c.  Composite of death or graft failure

  d.  Post-transplant diabetes mellitus

  e.  Serious adverse events (subdivided into respiratory, gastrointestinal, hepatobiliary, cardiovascular, musculoskeletal, skin, genitourinary, haematological, other transplant-related, other/unspecified)

  f.  Incidence of anaemia (defined as Hb < 13 g/dL in men, <12 g/dL in women) and severe anaemia (defined as Hb <11 g/dL in men, <10 g/dL in women) [at 3 and 6 months post-randomization only]

  g.  Incidence of leucopania (defined as total white cell count <3 x $10^9$ cells/mm$^3$) [at 3 and 6 months post-randomization only]

  h.  Incidence of neutropaenia (defined as neutrophil count <2 x $10^9$ cells/mm$^3$) and severe neutropaenia (defined as neutrophil count <1 x $10^9$ cells/mm$^3$) [at 3 and 6 months post-randomization only]

  i.  Incidence of thrombocytopaenia (defined as platelet count <75 x $10^9$ cells/mm$^3$) [at 3 and 6 months post-randomization only]

  j.  Levels of immunosuppressants at 3, 6, 12 and 18 months after randomisation

  k.  Proteinuria (compared by calculating ratio of geometric means) [at 3 and 6 months post-randomization only]. If albumin:creatinine ratio (ACR) value is available but not protein:creatinine ratio (PCR) for a given time-point, then it will be converted to PCR by the formula PCR = ACR x 1.6

  l.  Total cholesterol, HDL-cholesterol, non-HDL-cholesterol and triglyceride concentrations, and use of lipid-lowering therapy [at 6 months post-randomization only]

iv.   Tolerability

a. Reasons for stopping allocated treatment subdivided into suspected serious adverse reactions, serious adverse events (subdivided as in iii.e above), non-serious adverse reactions (including proteinuria, deteriorating graft function and other NSARs in same categories as SAEs) and other reasons (patient wishes or other)

These analyses will be conducted when all participants have completed 2 years and at a median follow-up of 5 years after transplantation. Participants will be censored at the date at their date of death or the date on which they were last known to be alive (based on flagging with the Office for National Statistics) once 2 (or 5) years have elapsed after the last patient was transplanted.

## 2.3 Tertiary comparisons

The primary outcome will also be assessed in the following subgroups (at 2 years after transplantation and after a median follow-up of 5 years):

- Induction therapy allocation
- Men and women
- Age (<60, ≥60 years)
- Deceased brain death, deceased cardiac death and living donors
- Categories of HLA mismatch defined by NHSBT allocation categories: 1 (no mismatches); 2 (0 DR and 0 or 1 B mismatches); 3 (0 DR and 2B, or 1 DR and 0 or 1 B mismatches); 4 (1 DR and 2B, or 2 DR)
- Sensitization status (high versus normal determined by calculated reaction frequency)
- First and subsequent transplants
- Categories of baseline graft function (eGFR <40, ≥40 <60 and ≥60 mL/min)
- Categories of baseline proteinuria (urine protein:creatinine ratio <30, ≥30 <50, ≥50 mg/mmol)

Exploratory analyses will also be undertaken of reported serious adverse events, non-serious adverse reactions and laboratory results (with due allowance made in their interpretation for the retrospective and exploratory nature of such analyses).

## 3   Details of analyses

### 3.1 Methods of analysis

The fundamental assessments of efficacy and safety will involve comparisons among all randomized patients in their originally allocated treatment group (i.e. "intention to treat" analyses).[1, 2] Analyses of categorical outcomes will be based on the first relevant unrefuted event of a particular type (i.e. either confirmed or not refuted during the adjudication process). All time-to-event analyses will be based on the first relevant event, and will use log-rank methods to calculate average event rate ratios (95% confidence intervals) and their associated 2-sided P-values.[2, 4] Comparisons of continuous outcomes will adjust for the baseline value if available, otherwise two-sample t-tests will be used.

The primary comparison between treatment groups of mean follow-up eGFR at 2 years post-transplantation (ie, on average 18 months post maintenance randomisation) will use analysis of covariance (ANCOVA), adjusted for the baseline value.[5] The baseline creatinine value used will be the latest value available prior to the maintenance randomisation (including a value from a sample

taken on the day of that randomisation). Any value within 6 months of the anniversary of the date of transplantation will be considered as that year's value. If multiple results are available within this window, the one closest to the anniversary date will be used.

## 3.2 Missing data

It is likely that values of creatinine will be missing at random (as they will be collected via the UK Transplant Registry and it is very unlikely that their randomized allocation would affect whether a hospital provides data to the registry). Baseline characteristics of those participants with available and those participants with missing data will be compared to identify any characteristics associated with missingness. The amount of missing data at various timepoints after transplantation will be displayed graphically.

Missing eGFR values will be imputed using a multiple imputation model, using baseline covariates of:

- Age (calendar year)
- Sex
- Type of transplant (living donor, donor after brain death and donor after cardiac death)
- Donor age (calendar age)
- Proteinuria at baseline (in categories: ≤30 mg/mmol, >30 mg/mmol)
- Available eGFR values (both before and after the value that requires imputation) of creatinine (hence eGFR) will be used as time-varying covariates (as exact values [to 1 decimal place]). Values before the maintenance randomisation will not be considered, nor any values on or after the date of failure of the transplant.
- Any characteristics associated with missingness

The imputation model will be stratified by the induction and maintenance therapy randomized allocations.

20 imputed datasets will be generated and then analysed using methods described in section 3.1.[6] These 20 results will then be combined in order to produce an overall result using standard procedures.[7] A complete case analysis will be conducted for comparison, but the primary emphasis will be placed on results after multiple imputation.

Any participant whose transplant has failed will be given an eGFR of 0 mL/min/1.73m$^2$ for all analyses after the date of failure. Sensitivity analyses will be conducted to vary this value of eGFR for failed transplants (range 0-10 mL/min/1.73m$^2$). Participants who have died will have eGFR values imputed as for other missing values (and a sensitivity analysis conducted with 0 imputed for any such participants and if such participants are excluded from the analysis).

Missing values of tacrolimus or sirolimus will also be imputed for any timepoint for which a value is expected but missing. The baseline variables used to impute missing values will include age, sex, type of transplant, donor age and any available values (as exact values to 1 decimal place) that occur after the maintenance randomisation but before failure of the transplant.

Missing values of urine protein:creatinine ratio will be imputed using baseline age, sex, type of transplant, randomized treatment allocations, available PCR values, angiotensin converting enzyme inhibitor or angiotensin receptor blocker use at baseline, diabetes status and cause of kidney

disease. Missing values for lipid fractions will be imputed using baseline age, sex, randomized treatment allocations, PCR at baseline, diabetes status and statin use at baseline.

All multiple imputation analyses will be implemented using SAS procedure MI. The expectation-maximization algorithm (which assumes a multivariate normal distribution) will be used to impute values (as eGFR is a continuous variable with an approximately normal distribution).

### 3.3 Allowance for multiplicity of comparisons

No allowance will be made for multiplicity testing in the primary comparison. For secondary and, particularly, tertiary comparisons, allowance in their interpretation will be made for multiple hypothesis testing,[1, 2] taking into account the nature of events (including timing, duration and severity) and evidence from other studies. In addition to the pre-specified comparisons, many other analyses will be performed with due allowance for their exploratory and, perhaps, data-dependent nature. Conventionally, two-sided P-values (2P) <0.05 are often described as "significant". But, the larger the number of events on which a comparison is based and the more extreme the P-value (or, analogously, the further the lower limit of the confidence interval is from zero) after any allowance has been made for the nature of the particular comparison (i.e. primary, secondary or tertiary; pre-specified or exploratory), the more reliable the comparison and, hence, the more definite any finding will be considered.

### 3.4 Tests for heterogeneity of effects

When a number of different subgroups are considered, chance alone may lead to there being no apparent effect in several subgroups in which the effect of treatment really is about the same as is observed overall. In such circumstances, "lack of direct evidence of benefit" is not good "evidence of lack of benefit", and clearly significant overall results would provide strong indirect evidence of benefit in some small subgroups where the results, considered in isolation, are not conventionally significant (or, even, perhaps, slightly adverse).[1, 2, 8] Hence, unless the proportional effect in some specific subcategory is clearly different from that observed overall, the effect in that subcategory is likely to be best estimated *indirectly* by applying the proportional effect observed among all patients in the trial to the absolute risk of the event observed among control patients in that category.[8] Tests for heterogeneity of the proportional effect on particular outcomes in specific subgroups will be used with allowance for multiple comparisons and for other differences between the subgroups to determine whether the effects in those subgroups are clearly different from the overall effect. If such subgroups can be arranged in some meaningful order then assessment of any trend in the proportional effects on outcome will be made.

### 4    References

1.      Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. Br J Cancer. 1976; **34**(6): 585-612.

2.      Peto R, Pike MC, Armitage P, Breslow NE, Cox DR, Howard SV, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. II. analysis and examples. Br J Cancer. 1977; **35**(1): 1-39.

3.      Racusen LC, Solez K, Colvin RB, Bonsib SM, Castro MC, Cavallo T, et al. The Banff 97 working classification of renal allograft pathology. Kidney Int. 1999; **55**(2): 713-23.

4.      Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. Prog Cardiovasc Dis. 1985; **27**(5): 335-71.

5.      Vickers AJ, Altman DG. Statistics notes: Analysing controlled trials with baseline and follow up measurements. BMJ. 2001; **323**(7321): 1123-4.

6.      Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009; **338**: b2393.

7.      Rubin D. Multiple imputation for non-response in surveys. New York: John Wiley; 1987.

8.      Collins R, MacMahon S. Reliable assessment of the effects of treatment on mortality and major morbidity, I: clinical trials. Lancet. 2001; **357**(9253): 373-80.