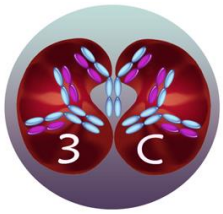

Campath, Calcineurin inhibitor reduction and Chronic allograft nephropathy

Data Analysis Plan for assessing clinical efficacy and safety of Campath-based induction 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 Campath-based induction therapy in the 3C Study. A separate Data Analysis Plan will describe the assessment of sirolimus-based maintenance 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 participants will then be followed by annual mailed questionnaires. In addition, all participants will be flagged with various NHS registries.

2 Comparisons of Campath-based versus basiliximab-based induction therapy

All comparisons will involve comparing outcome during the defined analysis period among all those participants allocated at first randomization to receive Campath-based induction versus all those allocated to receive basiliximab-based induction (i.e. "intention-to-treat" analyses).[1, 2] Any participants who are randomized but not subsequently transplanted will be censored on day 0.

2.1 Primary comparison

The primary comparison will be of the incidence of the first episode of biopsy-proven acute rejection (defined according to Banff 1997 criteria[3]) during the first 6 months after transplantation. Rejection episodes will be separated into acute cellular and acute humoral rejection and compared separately and in summary. Numbers of participants with rejection episodes of various severities will be shown, but statistical comparisons will not be made by severity. Participants who are subsequently randomised in the maintenance comparison will be censored on the day of their re-randomisation. Any participants not re-randomised will be censored 213 days (7 months) after transplantation.

2.2 Secondary comparisons

The secondary comparisons will be of the incidence of the first occurrence (during the first 6 months with censoring for maintenance randomisation as above) of:

i. All rejection episodes (biopsy-proven and presumed [suspected])
ii. Steroid-resistant rejection
iii. Delayed graft function (defined as any dialysis within the first 7 post-operative days)
iv. Serious infections (defined as infection requiring hospitalisation or an opportunistic infection). Comparisons will include:
    a. All serious infections
    b. CMV infection (subdivided into tissue invasive, non-tissue invasive infections and isolated CMV viraemia)
    c. BK nephritis and BK viraemia
v. Graft survival
vi. Participant survival
vii. Effects on full blood count:
    a. 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)
    b. Incidence of leucopaenia (defined as total white cell count $<3 \times 10^9$ cells/mm^3)
    c. Incidence of neutropaenia (defined as neutrophil count $<2 \times 10^9$ cells/mm^3) and severe neutropaenia (defined as neutrophil count $<1 \times 10^9$ cells/mm^3)
    d. Incidence of thrombocytopaenia (defined as platelet count $<75 \times 10^9$ cells/mm^3)
viii. Levels of tacrolimus at 1, 3 and 6 months

These analyses will be conducted as part of the safety analysis which will be performed when all participants have completed one year of follow-up after transplantation.

2.3 Tertiary comparisons

Long-term effects of the induction therapies will be explored as tertiary comparisons. Such comparisons are potentially biased by unequal numbers of participants undergoing the maintenance randomisation in the two induction therapy arms and therefore such analyses should be considered exploratory. Allowance will be made for multiple hypothesis testing, the nature and number of events and evidence from other studies. These comparisons will be made once all participants have completed 2 years follow-up after transplantation and again after a median follow-up of 5 years.

The following outcomes will be compared:

i. Graft outcomes:
    a. Graft rejection
        i. Biopsy-proven acute rejection (cellular, humoral and all)
        ii. All rejection (presumed and biopsy-proven)
    b. Graft survival
    c. Graft function (eGFR calculated using the 4 variable MDRD formula to compare latest eGFR available in all participants at time of analysis. Imputation rules for missing eGFR values are detailed in the maintenance therapy DAP.)
ii. Safety outcomes:
    a. Serious infections
        i. All serious infections (as defined in the protocol ie, all infections requiring admission to hospital or opportunistic infections)

ii. CMV infection (subdivided into tissue invasive and non-tissue invasive infections)
iii. BK nephritis

b. Cancer
   i. All cancer
   ii. Post-transplant lymphoproliferative disorder (in all participants and among EBV positive [including unknown] and negative serostatus at transplantation separately)
   iii. Non-melanoma skin cancer
   iv. Other cancer (ie, not ii or iii)

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. Post-transplant diabetes mellitus
   c. Levels of tacrolimus at 1,3, 6 months in the two arms. Levels of tacrolimus and sirolimus at 9 and 12 months, and then annually will be explored.

Exploratory safety analyses will also be undertaken of many other reported serious adverse events and non-serious adverse reactions (with due allowance made in their interpretation for the retrospective and exploratory nature of such analyses). In addition, the primary outcome will also be assessed in the following subgroups (at 6 months and 2 years after transplantation and after a median follow-up of 5 years):

- 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)
- Sensitization status (high versus normal determined by calculated reaction frequency > or ≤85%)
- First and subsequent transplants

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 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 Cox regression methods to estimate average event rate ratios (95% confidence intervals). The log-rank statistic will be used to calculate the associated 2-sided P-values.[2, 4] Comparisons of continuous variables (e.g. immunosuppressant concentrations) will use standard two sample t tests (after appropriate transformation if required). For outcomes that do not require time-to-event analysis (eg, delayed graft function), odds ratios will be calculated.

3.2 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 supplementary 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.

3.3 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, 5] Hence, unless the proportional effect in some specific subcategory is clearly qualitatively 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.[5] 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. Cox DR. Regression Models and Life-Tables. Journal of the Royal Statistical Society Series B (Methodological). 1972; **34**(2): 187-220.

5. 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.