Reviewer's report

Title: Non-AIDS defining cancers in the D:A:D Study - time trends and predictors of survival: a cohort study

Version: 1 Date: 1 July 2013

Reviewer: Francesco Donato

Reviewer's report:

Major Compulsory Revisions

1. The characteristics of the subjects at baseline (as shown in Table 2) for the overall cohort should be reported in a new Table 1, including two columns for the patients with and without NADC for comparison (statistical tests could add some information too).

2. Please define the number of patients lost during the follow up.

3. Please change “mortality rate” into “cumulative mortality” or similar phrase in all manuscript, because the findings describe as “mortality rate” really are not rates but proportions.

4. In the Materials and Methods section (page 7), the methodologies concerning the univariate and multivariate analyses performed should detailed. In particular, variables analyzed in the univariate model and maintained into the multivariate model should be mentioned. Specify whether the assumptions of proportionality of the Cox model have been verified. Moreover, in Table 3 a variable should be included, indicating whether a cancer (ADC or NADC) developed after the first NADC diagnosis. Why didn’t you include age as independent variable in the multivariate model, since age is significantly associated with mortality in the univariate model and it may be a confounder of the other predictors, regardless of the results of statistical significance test. There is a discrepancy between the data in Table 3 and the Results section. Indeed, you reported the following sentence “Risk factors for poorer survival after anal cancer were a previous NADC, HCV co-infection and disseminated cancer at time of diagnosis”. whereas Table 3 reported a non-significant association between the disseminated stage of disease and mortality [HR 2.49 (95% confidence interval 0.73-8.54, p = 0.15)].

Minor Essential Revisions

1. In order to improve the abstract the following sentence (Abstract, page 3) “The incidence of NADC / 1000PY ... and 2009/2010 respectively” is redundant and could be eliminated, otherwise data on the cumulative mortality at 1 and 5 years for NADC, anal cancer, lung cancer and Hodgkin's lymphoma should be added.

2. Why has the follow-up been interrupted on February 2010?
3. (Methods, page 6) “The present analysis is based on data collected up to 1st February 2010. By that date, information had been submitted on 1993 possible NADC; after excluding events occurring prior to baseline (the latest of 1 January 2004 or entry in the D:A:D Study) and rejected events (AIDS-defining cancers, dysplasias, subsequent events following a first event, in situ cancers particularly for the anal and head/neck regions), 880 NADC were ultimately included in analyses”. The sentence highlighted in yellow is not clear, since AIDS defining cancer and dysplasias are not NADC, and therefore they should not be included in 1993 possible NADC. Furthermore, the term “possible” is ambiguous, given the NADC definition reported in previous section (“Possible: based on a precise clinical description of the case, and where treatment had been initiated to support the description of an invasive NADC, but lacking supportive histopathological findings;”).

4. The method of calculation of the confidence intervals of the age-adjusted incidence rates is not specified in Materials and Methods.

5. In Table 2 and in the Results section a statistical test for comparison among the three specific NADC could be informative. In the same Table, I would suggest putting closer the column N and column (%), removing the symbol (%) in the heading, since the content of the parenthesis is already defined in any line. In the smoking status and in stage disease variables, the category “unknown” should be added. "Any use of ARV" should be changed into "Any use of cART".

6. (Results, page 8) "Also shown in Table 1 is the proportion of patients that developed an ADC (n=621) over followup." This sentence could be changed into "Also shown in Table 1 is the number of patients that developed an ADC (n=621) and type of ADC over followup.".

7. In the “Incidence over time” section of the results (page 8) a test for statistical significance could be added, specifying the type of test. Moreover, regarding the following sentence "Thus, the incidence of a first substantially NADC did not vary over the study period (p = 0.91 after age adjustment)" it is necessary to clarify the test used for the p-value calculation. There are a lot of numbers in the text, that makes reading difficult. Therefore, I would suggest taking off all the 95% CIs from parentheses.

8. It is necessary to specify that the mortality analysis consisted of mortality for all-causes in section "Mortality following NADC diagnosis" (page 9).

9. As regards "Mortality following NADC diagnosis" section (page 9), the data related to cumulative mortality of lung cancer at 3 years don’t meet with the data in Figure 2 where they are truncated at 2.5 years.

10. The following sentence (page 10) is misleading and I suggest changing it from "However, in multivariable analyzes, HCV co-infection and disseminated cancer at time of diagnosis were predictors of lung cancer mortality" to "However, in multivariable analyzes, HCV co-infection and disseminated cancer at time of diagnosis were predictors of mortality among patients with lung cancer ". 
11. The title of Table 3 must specify whether it reports univariate or multivariate analysis and the model used for the analysis.

12. The heading of Figure 2 should be rewritten as follows (or similar phrase): "Cumulative mortality NADC following diagnosis, stratified by type of NADC" (page 26).

13. I would suggest changing the term “race” into “ethnicity”.

Discretionary Revisions

1. The age-adjusted incidence rate could be calculated by using the population at the median cohort year (2007) or the European population as the standard truncated at 65 years.

2. A sensitivity analysis could be added. Indeed the study included also patients who had already had a NADC prior to 2004. The presence of a previous NADC could have an influence both on occurrence of a subsequent NADC and on survival. The authors could provide an analysis limited to those patients who did not have cancer prior to 2004 and verifying that the results are not substantially different from those presented in the article.

**Level of interest:** An article of importance in its field

**Quality of written English:** Acceptable

**Statistical review:** Yes, and I have assessed the statistics in my report.

**Declaration of competing interests:**

I declare that I have no competing interests