Reviewer's report

Title: People with diagnosed HIV infection not attending for specialist clinical care: UK national review

Version: 2 Date: 20 April 2015

Reviewer: Alison Howarth

Reviewer's report:

Discretionary Revisions (which are recommendations for improvement but which the author can choose to ignore)

It is interesting to hear the evidence provided for having left the country (line 102-06). Some mention of the evidence or lack of evidence for the remaining 163 cases would also be useful.

Ordering the variables in Tables 1 and 2 in the same way would help the reader.

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

Reference should be made to the data in Table 1 for the 456 individuals not receiving care in the UK (line 97).

Major Compulsory Revisions (which the author must respond to before a decision on publication can be reached)

It would be helpful to include more detail about the datasets and linkage (lines 25-32) to clarify that the cohort was created by linking records from the annual SOPHID surveys using full and part-linked identifiers (sex, date of birth, soundex code, and postcode of residence), as described in the Rice et al paper, and similarly how records were linked to reported deaths.

An explanation as to why and how newly diagnosed people were assigned to the reporting clinic or to the nearest clinic (lines 36-38) is needed.

Does “clinical record of attendance” (line 44) mean 1+ clinic visit for each of the years listed? And what does “patient location” (line 45) mean?

Please explain why the extended data were only collected for a small randomly selected sample of cases (lines 45-50). Rather than a vague description of all the data that were collected for the extended dataset, it would be helpful to have a clear definition of the variables that were used in the analysis – for example, last recorded CD4 count 2011. However, it would also be informative to define setting of HIV diagnosis, year of arrival in the UK, and social, psychological and economic issues and include them in the description of non-attenders.
A range of univariate statistical analyses are reported as being used (lines 71-74) but only findings from the logistic regression analyses are reported in Table 2 and the results section.

The statement that “proportions were presented where missing information was included” (line 76) is not clear – please re-phrase.

There is a difference in the way that non-attenders were identified in Scotland. Would it not, therefore, be more accurate to say (lines 86-92) that PHE provided clinics in E, W, NI with details of 3,452 individuals with no linked surveillance report in 2011 and questionnaires were completed for an additional 58 patients identified by clinics in Scotland and believed not to have received care in 2011? Resulting in a total of 3,510 individuals in the UK apparently not seen for care in 2011 / for whom the audit questionnaires were completed. And refer to the denominator in this (or a similar) way throughout the paper.

If basic data (age, gender, ethnicity) are available from SOPHID for the 3,452 individuals with no linked surveillance data, these individuals should be compared to the sample of 2,197 for whom the audit questionnaires were completed to give a measure of bias.

It is not clear where the cases come from in Table 2. Please provide the number of cases in the univariate analysis column, and clarify who the 308 cases and 136 cases are in the next two columns. Please also specify which variables were included in the multivariable analyses.

Please include some discussion of the response rate (proportion of individuals for whom audit forms were completed) as a further limitation in the conclusions.

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

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.