Reviewer’s report

Title: Profiling of individuals responsible for local HIV-1 transmission

Version: 1 Date: 21 April 2010

Reviewer: Janet Young

Reviewer’s report:

The manuscript “Profiling of individuals responsible for local HIV-1 transmission” by Kristen Chalmet et al (BMC article) addresses a very important issue – how to track and understand how transmission of HIV-1 takes place, and what differences might exist between route of transmission, the presence of co-infections, gender distribution, and origin. This phylogenetics data combined with demographic information, laboratory and clinical data appears to confirm previous epidemiological trends, and provides further detail concerning molecular events.

1. The number of individuals used for the study is appropriate for the general observations.

2. In the cluster observations and interpretation (Results, Transmission clusters - 180-205) it is appropriate to do an analysis without the large cluster of 57 as this revealed a difference in the association of clusters with Chlamydia infection, CD4 cell counts and presentation during acute infection. But this raises another question for this reviewer: could these associations be related to a biological phenomenon within an individual cluster, such as a specific viral variant, and not related to the transmission of the virus among a specific group of patients? Thus, what one might initially interpret as being related to the clustering phenomenon could be related to similarities between the particular virus variant being transmitted among individuals in the different clusters. It would be important to see if individual clusters show variation in the laboratory and clinical data that is a result of the subtype B variant that is simply averaged out when clusters are combined. An examination of some of the individual clusters might begin to show the underlying biological cause for differences between cluster and non-cluster transmission. (This is a major compulsory revision.)

3. In the Statistical Analysis section (Methods, last paragraph-149-158) of the manuscript, the authors state that in order to deal with confounding, they chose to conduct a backward stepwise procedure and keep in the model any variables with a moderate relationship (p <0.2) with the dependent variable. The stepwise procedure is an iterative process in which the software of choice is allowed to evaluate a set of variables for inclusion or exclusion in a model often based upon a sequence of f-tests and p values. Since these tests are carried out on the same data, there may be problems of multiple comparisons and a need for correction. Also, the p values are difficult to interpret since each is conditional upon previous runs of the regression.
Finally, evaluation of confounding is not conducted by evaluating p values and statistical significant relationships but upon evaluation of changes in parameter values of each variable in a model (with and without the variable) considering a priori described relationships. Confounding is a function of the complex interrelationships between exposures and the outcome of interest. During modeling to evaluate confounding, one would look for a change in parameter estimate for each variable considered as a confounder. A change in beta of 10-15% is substantial and indicates a subsequent need for adjustment/inclusion in the model to control for confounding.

(This is a minor essential revision. Authors could, for example, decide to revise the sentence before and just say they were selecting the best significant predictive set of variables.)

**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.