Reviewer's report

Title: Human papillomavirus (HPV) types 16, 18, 31, 45 DNA loads and HPV-16 integration in persistent and transient infections in young women.

Version: 2 Date: 20 July 2010

Reviewer: Marc Baay

Reviewer's report:

HPV types DNA loads and HPV-16 integration in persistent and transient infections in young women.
Ramanakumar et al.

This paper is well-written, with a sound methodology, a clear presentation of the results, and adequately discussed. It can therefore be accepted for publication after minor revision, since I do have some remarks.

In the HPV-16 integration section, last sentence of the first paragraph, the authors discuss HPV 6 E6/E2, rather than HPV 16 E6/E2.

In the last paragraph of the results section, the authors go a long way to describe how they have ensured that they were able to measure E2 and E6 quantitatively. I was wondering however, if it would not be possible to have integration in the presence of episomal HPV DNA (as has been described in the literature). Since these young women frequently have high viral loads, the effect of integration on E6/E2 ratio might easily be obscured. Maybe the authors can comment on this.

In the discussion, page 12, last paragraph there is clearly a chicken or egg paradox. Does disruption of E2 lead to integration, or does integration lead to disruption of E2?

On the next page, the authors quote Brown et al [46], on disruption of E2 in 25% of infections, however, the paper in the ref list under 46 does not discuss HPV integration, surely there must have been a mix up.

In refs 5 and 35, it is not Snidjers but Snijders, like in ref 40. Update ref 39, since this paper has been published now.

It took me quite a while to understand figure 2. This could of course be a personal problem, but I would like to suggest helping readers by giving an example, for instance: upper left box, viral load at entry versus viral load at 6 months, lower left box, viral load at entry versus viral load at 24 months. This way, it will be easier to understand. Furthermore, it would be helpful to give the correlation coefficient for each box (e.g., in the upper left corner, except for 18 vs 24).

Again, figure 3 puzzled me. In the text, page 9, first sentence I read: ” infection
clearance was inversely associated with viral load" [high viral load, low clearance]. In table 3 I read: “third viral load tertile (highest) is taken as reference”, and indeed values for second and first tertiles are higher [suggesting higher clearance rate]. In the legend to the figure (incidentally, where you use lower, middle and upper tertile, why not use the same terms in table and figure consistently) I read that the upper tertile (= third? = highest?) is the solid line. All women start with an infection [1.00], and than the line goes down every time a woman clears the infection. But in that case, shouldn’t the solid line end up on top and not as the lowest? I am afraid I’m lost and need some help.

**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.

**Declaration of competing interests:**

'I declare that I have no competing interests'