Reviewer's report

Title: A coarsened multinomial regression model for perinatal mother to child transmission of HIV

Version: 1 Date: 20 February 2008

Reviewer: Laurence S. Magder

Reviewer's report:

General

In this generally well-written paper, the authors propose a relatively simple model for estimating in utero and perinatal transmission of HIV in the situation when, due to incomplete data, there is uncertainty regarding the timing of HIV acquisition for some of the children.

The strength of the paper is that the described method is relatively simple and takes into account a problem commonly confronted (and often ignored) in PMTCT studies, as well as in other studies of disease incidence. Also, their recognition that there are different risk factors for in utero infection and intrapartum infection is important and sometimes ignored.

In fact, I used the methods that the authors are proposing (for the case of two time intervals) in a paper published a few years ago (Risk factors for in utero and intrapartum transmission of HIV#, J Acquir Immune Defic Syndr 2005, 38:87-95). However, in that paper I did not describe the method in detail, and it would be good to have a description of the method in the literature. Also, their approach is a bit more general.

My main problem with this paper is that the authors propose models for the probabilities of having a first positive HIV result in various intervals rather than models for probabilities of real clinical and scientific interest: i.e, the probability of in utero transmission or intrapartum transmission (which are denoted as A1, and A3 in the paper). In doing so, they cannot accommodate misclassification in their model. This is discussed in greater detail below.

Discretionary Revisions.

1. Instead, of modeling the probabilities of having a first positive HIV result in various intervals, the authors should consider modeling the probability of various modes of transmission (in utero, intrapartum, breast milk). The latter are of greater scientific interest (i.e., we want to know probabilities of, and risk factors for transmission, not for positivity), as the authors acknowledge. Also, the probabilities of transmission are more likely to be directly related to covariates than the probabilities of first positivity. For example, the probability of intrapartum transmission is likely influenced by predictors such as the duration of rupture of membrane, complicated delivery, viral load at the time of delivery, etc., whereas
the probability of having a positive HIV test is related to those predictors only through transmission.

To make this change, the authors need to specify an additional latent random variable which indicates whether the child acquired HIV in utero, intrapartum, through breast milk, or remained uninfected. Then the authors need to specify a model for the probability of HIV positivity at each time interval given each possible type of transmission. At this step, if they specify that the probability of positivity at birth is 1 if the child was infected in utero (and 0 if the child was infected intrapartum or uninfected), then this more complex model reduces to the model that they have proposed. This approach is much cleaner and more general, and includes their model as a special case, while explicitly incorporating assumptions about the relationship between mode of transmission and timing of positivity. This is the approach I used in my paper (mentioned above).

Minor essential revisions.

2. In the second paragraph describing the simulation study, the authors say “We fit cumulative and conditional regression models using the proposed coarsened multinomial (CM) regression models and logistic regression models.” What link was used for the CM models? (If they said, I missed it). Also, if they used a logit link, then the CM models are, in some sense, logistic regression models too, so that this sentence is a bit confusing. I suggest that they reword.

3. In the description of the HPTN 024 analysis, the authors say “In this analysis, we examined the association between PMTCT and antibiotics as taken as opposed to as randomized, comparing outcomes for infants born to mothers randomized to antibiotics who delivered prior to termination of the study drug to infants born to mothers randomized to placebo or who delivered after termination of the study drug.” Two problems: 1) the first part says that they are doing an “as taken” analysis, but the second part seems to contradict this saying they are comparing those randomized to one group to those randomized to another group. 2) The expression “prior to termination of the study drug” was initially confusing to me. I first read it as though it was referring to a women’s use of the drug, but now I think I see that it means prior to the time the investigators stopped administering the drug to patients. A re-write would make this clearer.

4. In reference to the Balasubramanian and Lagakos paper, the authors say “Because we examined a breastfeeding population and are interested in categorizing infection timing, this approach is not suitable.” This is not a valid point. First, B and L were also interested in timing, and second the author’s model itself does not account for breastfeeding transmission either.

5. Similarly, in the second from last paragraph the authors say that their model is valid for breastfeeding populations. This true only because their model pools breastfeeding transmission with intrapartum transmission. Since the risk factors for intrapartum transmission are going to be very different from breast-feeding transmission, pooling the two groups is not scientifically helpful. Thus, I think it is
fairer to say that their model does not really account for breastfeeding transmission.

6. I found Appendix A confusing. J=4, so both Y and Y* should be 5x1 vectors. But they are shown as 2x1 and 3x1 vectors. Also, the later three time intervals in the picture seem irrelevant to the text in the appendix. In any event, I am not sure that Appendix A is needed. I think a few examples in the body of the paper described in prose might be all you need to help readers understand how the Ys are constructed.

7. In Appendix B they say "We calculated the probabilities of in utero infection, ï¬†1ï¬†ï¬†. This sentence incorrectly equates the probabilities of transmission with parameters that stand for the probabilities of first positivities. The ï¬†ï¬†s in this sentence should be replaced with A1, A2, and A3. This error illustrates the main problem with their approach. They would like to equate the two sets of probabilities, but really canât.

What next?: Accept after discretionary revisions

Level of interest: An article whose findings are important to those with closely related research interests

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