Author's response to reviews

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

Authors:

Charlotte C Gard (gardc@u.washington.edu)
Elizabeth R Brown (elizab@u.washington.edu)

Version: 2 Date: 1 May 2008

Author's response to reviews: see over
Author’s response to reviews

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

Authors:

Charlotte C Gard (gardc@u.washington.edu)
Elizabeth R Brown (elizab@u.washington.edu)

Version: 2  Date: 1 May 2008

Author’s response to reviews: see over
We thank the reviewer for his helpful comments. We have modified the manuscript to address each of his concerns, as described below.

---

REVIEWER: Laurence S. Magder

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).

Our primary purpose in this manuscript is to address the problem of incomplete data, which is generally ignored in perinatal mother to child transmission (PMTCT) analyses and was of particular concern in the HPTN 024 study. We note that, as PMTCT analyses do not routinely adjust for sensitivity, the interpretation offered by our approach is consistent with that of approaches commonly used in the PMTCT literature. That said, we believe the reviewer makes an important point, and we have given considerable thought as to how we might address it.

Our approach currently relates an infant’s coarsened test result, $Y_i$, to his or her complete (unobserved) test result, $Y_i^*$. We could introduce a second latent variable, $\hat{Y}_i$, to represent an infant’s true infection status, as the reviewer suggests. $Y_i^*$, then, represents an imperfectly measured version of $\hat{Y}_i$.

To estimate associations between covariates and infection status, we would need to link the probability of infection at each time point to each possible coarsened test result, $Y_i$. This would require two models, the first relating $\hat{Y}_i$ to $Y_i^*$ and the second relating $Y_i^*$ to $Y_i$. In relating $\hat{Y}_i$ to $Y_i^*$, we could incorporate information about the sensitivity of testing in the manner of Magder and Hughes (American Journal of Epidemiology. 1997;146:195-203).
Two features complicate implementation of such an approach. First, our outcome is multinomial in nature. Thus, where Magder and Hughes expand the data to treat each subject as both diseased and not diseased, we would need to expand the data to treat each infant as infected in utero, infected during delivery, and infected due to early breastfeeding. Second, because we do not observe an infant’s complete test result $Y_i^*$ (as do Magder and Hughes) but, instead, observe a coarsened version of $Y_i^*$, we would need to weight each observation in the expanded dataset according to the probability of infection given the coarsened test result. To calculate the necessary conditional probabilities is non-trivial, as there are as many as six possible coarsened test results for a given $Y_i$ (see table below depicting possible combinations of $\hat{Y}_i, Y_i^*$, and $Y_i$).

<table>
<thead>
<tr>
<th>$\hat{Y}_i$</th>
<th>$Y_i^*$</th>
<th>$Y_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1 0 0)</td>
<td>(1 0 0)</td>
<td>(1 0 0)</td>
</tr>
<tr>
<td></td>
<td>(1 1 0)</td>
<td>(1 1 1)</td>
</tr>
<tr>
<td>(0 1 0)</td>
<td>(0 1 0)</td>
<td>(0 1 1)</td>
</tr>
<tr>
<td></td>
<td>(1 1 0)</td>
<td>(1 1 1)</td>
</tr>
<tr>
<td>(0 0 1)</td>
<td>(0 0 1)</td>
<td>(0 1 1)</td>
</tr>
<tr>
<td></td>
<td>(1 1 1)</td>
<td></td>
</tr>
<tr>
<td>(0 1 0)</td>
<td>(0 1 0)</td>
<td>(0 1 1)</td>
</tr>
<tr>
<td></td>
<td>(1 1 0)</td>
<td>(1 1 1)</td>
</tr>
<tr>
<td>(0 0 1)</td>
<td>(0 0 1)</td>
<td>(0 1 1)</td>
</tr>
<tr>
<td></td>
<td>(1 1 1)</td>
<td></td>
</tr>
<tr>
<td>(0 0 1)</td>
<td>(0 0 1)</td>
<td>(0 1 1)</td>
</tr>
<tr>
<td></td>
<td>(1 1 1)</td>
<td></td>
</tr>
</tbody>
</table>

To incorporate this modification into the current framework would add considerable complexity. Not only does it introduce a second latent variable but, unlike the current approach, it requires that an EM-like algorithm be used for estimation. Given that the simplicity of the current approach is one of its strengths (as noted by the reviewer), we believe it preferable to present the current approach as described and address this extension in future work. We have added the following paragraph to the discussion to speak to this issue.

“While we have attempted to assess the impact of misclassification in our simulations, our approach uses probabilities of first testing positive to estimate transmission probabilities and, in doing so, does not formally account for misclassification due to the imperfect sensitivity of testing. A possible extension of this approach would involve introduction of a latent variable representing
an infant’s true infection status. One could link an infant’s probability of infection at a given time point to his or her coarsened test result via his or her complete (unobserved) test result and, in doing so, incorporate information about the sensitivity of testing in the manner of Magder and Hughes (American Journal of Epidemiology. 1997;146:195-203). Given the multivariate nature of the outcome and missingness in the test result, such an extension would introduce considerable complexity. Further, unlike the current approach, it would require use of an EM-type algorithm.

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.

This sentence has been re-written to read “We fit the cumulative and conditional regression models using standard logistic regression and the proposed coarsened multinomial (CM) regression models with the logit link.”

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.

The following sentence has been added to the second paragraph of the HPTN 024 subsection: “Administration of the study drug was halted on March 5, 2003.” The sentence in question has been re-written to read: “In this analysis, we examined the association between PMTCT and antibiotics, comparing outcomes for infants born to mothers randomized to antibiotics who delivered prior to March 5, 2003 to infants born to mothers randomized to placebo or to mothers randomized to antibiotics who delivered after March 5, 2003.”

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.

The final two sentences of this paragraph have been replaced by the following: “Balasubramanian and Lagakos [11] provide methods for estimating the continuous distribution of the timing of in utero and peripartum transmission that account for the imperfect sensitivity of the HIV assay. The authors developed the approach for settings in which there is no risk for infection following birth and, therefore, do not address the potential impact of breastfeeding.”
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.

We have modified this sentence to read “While the model can be used in a breastfeeding population, it does not allow us to separate intrapartum transmission from early transmission due to breastfeeding.” We note that, in our simulations, we allowed for additional positive test results at the 4 to 8 week visit due to breastfeeding. Our model performed well when compared to standard logistic models despite this misclassification.

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 Y’s are constructed.

We have moved examples A and B, with corrected Y*, into the body of the manuscript (Section: Methods, Subsection: The coarsened multinomial model, Paragraph 4) and have removed Appendix A.

7. In Appendix B they say “We calculated the probabilities of in utero infection, \( \pi_{i1} \).” This sentence incorrectly equates the probabilities of transmission with parameters that stand for the probabilities of first positivities. The \( \pi_i \)'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 can't.

We have changed our notation to use A1, A2, and A3 as opposed to \( \pi_{i1}, \pi_{i1}+, \pi_{i2}, \) and \( \pi_{i2|1-} \). References to equations (6), (7), and (10) in paragraph 2 have been replaced by regression equations for A1, A2, and A3. In addition, we have changed our descriptions of A1, A2, and A3 (Section: Introduction) to refer to transmission probabilities rather than transmission rates.

For further clarification, we have also made the following changes:

- Page 9, Section: Results, Subsection: Simulation study, Paragraph 1. Last sentence has been modified to read “Relative bias (not shown in table) of estimators of treatment effect on perinatal and intrapartum transmission ranged from 0.007 (conditional CM model) to 0.711 (cumulative logistic model), corresponding to the scenario where treatment was assumed to increase the risk of in utero transmission but decrease the risk of intrapartum and overall perinatal transmission (TE4) and the visit process resulted in the most missing data (VP3).”

- A sub-subsection relating to maximum likelihood estimation of parameters has been added (within Section: Methods, Subsection: The coarsened multinomial model) and the final two paragraphs of the Modeling cumulative rates of transmission sub-subsection moved to this sub-subsection.