

Homsy J, Bunnell R, Moore D, King R, Malamba S, et al. Reproductive Intentions and Outcomes among Women on Antiretroviral Therapy in Rural Uganda: A Prospective Cohort Study. *PLoS ONE* 2009; 4(1): e4149. doi:10.1371/journal.pone.0004149.

This study investigated desire for more children, birth control use, and pregnancy incidence among 733 HIV-infected women in Uganda for two years after they initiated anti-retroviral therapy.

Problem 1 (negative conclusions based on p-values). There appear to be two instances of this problem. At the end of the Results, several predictors are described as “no longer independently associated with pregnancy” in the adjusted model. This is contradicted by the estimates. Here are the univariate and adjusted estimates and CI’s:

Variable	Univariate HR	Adjusted HR
CD4 count	1.001 (1.000-1.002)	1.025 (0.98-1.07)
Death of child	1.63 (1.06-2.50)	1.94 (0.64-5.95)
Death of spouse	0.19 (0.04-0.88)	0.49 (0.13-1.83)
Length of relationship	0.97 (0.94-0.99)	0.82 (0.65-1.95)

All of the variables (except having experienced the death of a spouse) have adjusted estimates that are further away from 1.0 (no effect) than their univariate estimates. We can gain some additional insight by examining how the SE’s (of the log HR’s) changed between univariate and adjusted.

For CD4, the SE increased 44-fold, which suggests that the univariate result is actually per 1 cell rather than per 50 as labeled. If so, we cannot see how the univariate and adjusted estimates and CI’s really compare, because the precision given for the univariate results is inadequate. Severe collinearity with some other variable in the model could possibly produce such a huge increase in the SE, but it would have to be very severe and I see no obvious candidates for which other variable would be highly correlated with CD4 count.

The SE for death of child has increased by 2.6-fold, which could be due to collinearity. This greater uncertainty is not the same as certainty that there is no longer an association, as the phrasing in the paper implied. Deleting other variables, one by one, from the multivariate model might have revealed the source of the collinearity.

The SE for death of spouse is almost the same (0.9-fold). The change in the estimate, closer to no association, could be due to confounding. Even though 0.49 is closer to no association, it is still very far from 1.0.

The SE for length of relationship has increased 21-fold. This again suggests a scaling problem, or possibly severe collinearity.

Another possible instance of Problem 1 is the claim near the beginning of the Discussion that “neither [women’s nor partners’ desire for children] was significantly associated with pregnancy.” These had upper confidence bounds of HR=3.4 and HR=2.8, which presumably are large enough to be “significant”.

Problem 2 (vague, misleading phrasing). Near the end of page 6 is the statement: “no association was found between pregnancy and the number of children the women had.” A comparison with other studies should focus on the estimates and CI’s and provide a statement about whether the results support or argue against the findings of the previous studies and how strong the evidence is. The reference category of no children in Table 3 included very few women, so the results shown are fairly uninformative.

Problem 7 (entangled outcomes and predictors). This was avoided by using time-dependent covariates. At the top of page 4, they note, “At a given time point, the value of the predictor variable was determined by the most recent prior measurement.”

Best Practice 1 (directly relevant estimates and CI’s). Table 3 shows hazard ratios that directly quantify the association of the variables with pregnancy, and CI’s are also shown. Many of the main conclusions at the beginning of the discussion concern results from Tables 1 and 2 and Figure 2, and those do not include any CI’s. In addition, Table 1 is potentially confusing because it does not directly quantify associations with pregnancy, but it addresses them indirectly by providing separate summaries and p-values comparing women who did and did not become pregnant. Its results concerning body mass index, CD4 count, length of relationship, and condom use may appear to contradict the more direct univariate results in Table 3. Summarizing the entire cohort together may have been a better choice for Table 1. In general, a standard Table 1 can compare groups defined by a key predictor variable (such as treatment arm), but comparing groups defined by outcome is usually not a good idea.

Best Practice 2, 2a (interpretation based on estimates and CI’s). The interpretations at the beginning of the Discussion concern estimated rates and how high or low they are, which follows this best practice. Also having CI’s and factoring those in to the Discussion could have improved it by, for example, noting that not only did 93% of women not want more children, but the lower confidence bound on this rate was also quite high. I see little attention given to the actual estimates and CI’s in Table 3, even though they were prominently presented. See problem 1, above.

Best Practice 3 (discuss what may be true in general). The paper does seem to focus on implications for what may be true in general. Some of the instances of problem 1 may have resulted from focusing on p-values instead from this study instead of what the estimates implied about what may really be true.

Best Practice 4 (state what you did find or learn). The major conclusions are positive and supported by the results.

Best Practices 5, 5a (learn as much as you can). This seems to be a comprehensive paper, with multifaceted analyses addressing a number of related issues. ITT versus per-protocol analysis does not arise in this observational study.

Best Practices 6, 6a (include scientific considerations). The discussion appropriately utilizes some outside knowledge in its interpretations, notably findings from other studies and knowledge of general population rates for many of the variables examined (pages 6-7). They did not use formal multiple comparisons adjustments. Some p-values in Table 1 are from omnibus tests, but those are not key analyses in any case.

Best Practice 7 (prefer accuracy to conservatism). They do not appear to have introduced conservative bias in the main analyses.

Checking assumptions. There is mention of checking linearity assumptions for numeric predictors, but no mention of checking the proportional hazards assumption for Table 3.

Accounting for dependence in the data. GEE methods accomplished this.

Survival analysis details. There are considerable details of event ascertainment and confirmation, censoring times, and loss to followup (in Figure 1).

Missing data. There was very little missing data, so this concern did not apply.

Insufficient precision. See problem 1 concerning CD4. Otherwise, precision was generally adequate and not excessive.

Poorly scaled predictors. See problem 1 concerning CD4 and length of relationship. The adjusted body mass index results in Table 3 may be another instance, even though it is labeled as a dichotomous variable; it may really be per 1 unit of BMI.

Use of “significant” alone. This occurs in several places.