Both papers seems generally good to me, but it is still possible to find some instances of the common problems:

Heisler M, Faul JD, Hayward RA, Langa KM, Blaum C, Weir D. Mechanisms for Racial and Ethnic Disparities in Glycemic Control in Middle-aged and Older Americans in the Health and Retirement Study. *Arch Intern Med* 2007; **167**: 1853-1860.

This study investigated differences in HbA1c among self-characterized diabetics who were white, black, or Latino. It found worse mean values among black and Latino respondents than among white respondents, with little of the differences explainable with other measured variables.

Problem 1 (negative conclusions based on p-values). The initial result reported in the Abstract was: "There were no significant racial/ethnic differences in HbA1c levels in respondents not taking antihyperglycemic medications." This appears to be based on the p-values in Table 1. Because they use "significant" to mean "important" in many places throughout the paper, this statement could easily be read as implying that the differences in those not taking medication was too small to be important. The actual black-white difference in mean HbA1c was 7.17-6.39=0.78; this difference was not actually given in the paper, so of course no CI for it was given either. We can deduce the CI's based on the p-values given in Table 1: for not on medication we have 0.78 (-0.36 to 1.92) and for on medication we have 0.85 (0.34 to 1.36). This is a lot of overlap, and we can assume that the difference of 0.78 is large enough to be important because it is nearly as large the difference they do investigate (0.78, versus a 0.85 difference among those on medication) and actually larger than the adjusted difference (0.73) from their final multivariate model (in Table 3). For the black-white disparity, the main difference between those on vs off medication was therefore the larger p-value due to the small number of subjects not on medication. They therefore should have noted that the black-white disparity also appeared to exist among those not on medication, but the CI in that group was too wide to rule out the possibility of no difference. For Latino respondents, the contrast between those on and off medication is larger. We can deduce a p-value for the interaction of Latino vs white with medication status from the information they give, obtaining p=0.16. Thus, while there is a suggestion of a different white-Latino disparity by medication status, the evidence is not conclusive.

Fortunately, the authors did not attempt to draw any conclusions from the "lack" of a disparity among those not taking medication. But this does not mean that others will not mistakenly attempt to do so. They did note that multiple regression models that included all respondents produced similar results to the medication-only analyses. Their main conclusion in the Abstract does not distinguish those on and off medication.

Another possible instance of Problem 1 is on page 1858: "In contrast, socioeconomic status, access, and health care quality variables, except for health insurance status in respondents younger than 65 years, were no longer independently associated with HbA1c levels in the fully adjusted models." This is less clear, because the fully adjusted model in Table 3 does not show an age-group-specific estimate for health insurance status, and many of the variables mentioned did not have $p<0.05$ in the unadjusted model. So the results forming the basis of this statement are not clear. In Table 3, note that having had HbA1c measured in the last 12 months, presumably a quality indicator, has a larger estimated effect in the fully adjusted model than in the unadjusted.

Problem 2 (vague, misleading phrasing). In the discussion of limitations on p. 1859, they write concerning inaccurate self reporting and social desirability bias, "There is no evidence to suggest, however, that different racial/ethnic groups are more susceptible to such bias than others." We cannot tell whether this has been studied and racial/ethnic differences are known to be small, or if there is simply nothing known about the issue. Also, the blanket claim that "there is no evidence" would require an extensive literature review, and the lack of any mention of such a review or citations for this statement casts doubt on whether any such review was done. It probably would have been better for them to omit this sentence.

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).

At the end of the Results, several predictors are described as "no longer independently associated with pregnancy" in the adjusted model, and this is restated at the end of the first Discussion paragraph by noting the "only" variables that were independently associated. The implication that univariate associations have been explained away 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 farther away from 1.0 (no effect) than their univariate estimates. We can gain some additional insight by examining how the widths of the confidence intervals (on the log scale) have changed between univariate and adjusted results.

For CD4, the CI width increased nearly 50-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: the effect per 50 cells could be anything from 1.025 to 1.078. 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 CI width 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 CI width 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 CI width for length of relationship has increased 21-fold. This again suggests a scaling problem, maybe years in one model and months in the other.

If they had paid any attention to the estimates and CI's that related to the above statement, they would have found these apparent errors in their results. They then should have rephrased the statement to reflect how the estimates actually changed between the univariate and multivariate models and the CI's around the estimates in the multivariate models.

The last sentence of "Enrollment and Baseline Characteristics" section on page 4 states: "Pregnancy was not associated with the women's allocation to the various ART monitoring arms of the HBAC trial (Table 1)." An absolute statement like "was not associated" would only be true if pregnancy rates were identical in the different arms. The rates were similar, but not identical:

Pregnancy rates (calculated from tallies given in Table 1):
Clinical+CD4+VL: 40/235 (17.0%);
Clinical+CD4 43/248 (17.3%);
Clinical only 34/225 (15.1%).
So better phrasing would have been: "Pregnancy had very little association with …". Better still would have been including this variable in Table 3 so that estimates and CI's for the associations would be directly available.

Near the end of the first Discussion paragraph on page 4 is a statement that: "neither [woman's nor partner's desire for children] was significantly associated with pregnancy." This may have been meant to simply say that both had p>0.05, but it could easily be interpreted to mean that there was evidence against any important association. The estimates and CI's from Table 3 were:

Variable	Univariate HR
Wants more children	1.37 (0.55-3.41)
Partner wants more children	1.42 (0.72-2.78)

The estimates may be too large to be considered no important association. In addition, the upper confidence bounds are quite large, so any implication of strong evidence against an important association is not supported by the data.

Page 7-8: "this variable was not statistically associated with pregnancy in our study". This again may simply be stating that p>0.05, but it might be interpreted as implying that the observed difference is an illusion and the true rates must be identical. Better phrasing might have been "was not substantially associated", but the estimated association is probably too large support such a statement:

Variable	Univariate HR
Experienced physical abuse by partner in last 12 months	1.36 (0.91-2.02)

Note again that the upper confidence bound is large, so any implication of strong evidence against an important association is not warranted.

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." The reference category of no children in Table 3 included very few women, so the results shown are fairly uninformative. The results relevant for this statement are:

Variable		Univariate HR
# Live Children (vs 0)	1-2	1.61 (0.50-5.25)
	3+	1.26 (0.40-4.01)
# Dead Children (vs 0)	1-2	1.35 (0.86-2.12)
	3+	1.26 (0.74-2.16)

The estimates, and especially the confidence bounds, are too far from 1.0 (no association) to allow phrasing implying "no association" to be appropriate. This could also be considered an instance of Problem 1.

Explicitly avoided: 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."