Reviewer's report

Title: Herpes simplex virus type 2 seropositivity and relationship status among U.S. adults age 20 to 49: a population-based analysis

Version: 1 Date: 24 June 2010

Reviewer: Maya Sternberg

Reviewer's report:

Major Compulsory Revisions

This paper does identify a subpopulation of the US that is often ignored by the sexually transmitted disease community, which focuses more on young adults and racial disparities. In addition, this paper attempts to use a Nationally representative data set to parse out the excess risk associated with relationship status among adults 20-49 years of age, unfortunately it does not succeed at being very compelling or informative from a methodological standpoint. First of all, NHANES has no information about when the person acquired the HSV infection; it is unclear that their current relationship status has any bearing on HSV-2 seroprevalence (which is based on lifetime exposure). In general, NHANES, as a cross sectional study, is not well-suited to address the type of epidemiologic question.

I find it very hard to believe that total lifetime sex partners did not remain statistically significant in any of the models presented. Total lifetime partners tends to not increase linearly with lifetime number of partners, so it may be necessary to consider a log transformation or consider the way Fleming (or other authors) categorize number of sex partners. I personally suspect, that if you included lifetime sex partners in a way that acknowledges this underlying non-linear relationship with the log odds, that it would in fact be significant and most likely the relationship status variable may no longer be statistically significant. If this is the case, it undermines the intent of what this paper set out to show. To give readers confidence that in fact lifetime sex partners was not an important predictor or confounder in these models the authors need explain in more detail how this variable was treated in the model and possibly even present their final model with this variable included, because at a minimum this variable is sure to be a confounder as it is significantly related to both marital status and herpes.

Please define and justify more specifically what “large” effect size means when developing the model, this is an arbitrary term that has different meanings for continuous variables vs categorical variables. In addition, often large effects are accompanied with large standard errors, which means one is building a model based on highly variable point estimates, what is the justification for doing this? For example, the odds ratio reported in the results section of 8.05 for relationship status in the 20-29 year old model is based on a confidence interval of \(\exp(2.086 \pm 1.96 (1.19) = (0.78, 83.0)!\) With a confidence interval like this you may as well
The argument made to present two logistic regression models for 20-29 and 30-39 year olds is rather weak. Just because the underlying relationship of HSV-2 to age is non-linear does not make a strong argument to provide separate models for those below and above 30. If there were extensive age by other variable interactions that might make a better reason, in my opinion, to report different models by age, or substantive reasons determined in advance of looking at the data (such as biologic differences between the groups that imply different predictors need to be considered).

I question the authors’ interpretation of the relationship status variable in the logistic model for 20-29 year olds. The authors interpret the relationship status beta coefficient from the 20-29 model as an OR for married/cohabiting vs single persons among those aged 20 (8.05); this implies that the age variable for those who are 20 was recoded to zero when the model was fit—however, this is not clearly explained in the methods section nor in the table. If on the other hand, age was not recoded so that those who were 20 receive the value 0 etc, then the actual odds ratio for the relationship status variable at age 20, as shown in the table, would be \( \exp(2.0892-.0839(20)) = 1.3 \), which is not a large effect size and this would in fact mean that the large effect size of this variable is outside the age range of the domain analyzed for the model reported. The authors need to explain somewhere how age was coded to arrive at this interpretation from the model, currently the table suggests that age is coded in years and hence the interpretation of the beta coefficient is wrong.

In general, while I think the authors identify an area of opportunity for research and public health messaging for HSV-2, I do not think NHANES is suited to make this a compelling story and the fact that lifetime sex partners did not even make it into the model is so strange that I attempted to redo the model using NHANES 99-08 and I do find that total lifetime partners as either a log transformed continuous variable or a categorized variable (0, 1, 2,3-5, 6+) is highly significant Wald F p-value <0.0001. I highly recommend you double check your original finding that lifetime sex partners if not significant in these models.

Minor Essential Revisions

Please indicate the specific test name for the p-value reported in the logistic regression models, in complex survey data there are many different tests such as Wald Chi-Square, Wald F, Satterthwaite Adjusted statistics etc. SURVEYLOGISTIC default is the Wald Chi-square tests which are a bit more liberal and will tend to give smaller p-values than those adjusted for the degrees of freedom for the complex design of the NHANES survey- though in the case of combining 10 years of data the degrees of freedom will be large enough to make little difference between Wald Chi and Wald F (which is SUDAAN’s default test).

A nuance which arises by stratifying at 30 is the interpretation of the relationship status variable. Those who are 20-29 are probably mostly comprised of married or never married; whereas in the 30-39 age group one expects more of those who are not cohabitating to be from the other groups divorced/separated/widowed. Have the authors considered what the differences
between divorced/separated/widowed and never married are in 30-49 year olds—are they similar enough to be combined into one group?

Discretionary Revisions

Possibly Create separate graphs for self reported genital herpes and HSV-2 seroprevalence – the current graphs are too dense and lose any value in telling the story- too many lines makes it confusing

The authors use 10-years of NHANES data (and I assume they recalculated the weights as described by NCHS when combining cycles); one question is whether combining the 10 years is valid in so much whether sexual behavior has not changed over the past decade, HSV-2 seroprevalence etc. May want to address this in the Discussion

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'