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

Authors:

Greta R Bauer (greta.bauer@schulich.uwo.ca)
Nooshin Khobzi (nooshin.khobzi@schulich.uwo.ca)
Todd A Coleman (todd.coleman@schulich.uwo.ca)

Version: 2 Date: 6 October 2010

Author’s response to reviews: see over
BMC Infectious Diseases Editor and Reviewers:

We would like to thank the reviewers for their thoughtful reviews of our manuscript. Responses to specific comments are outlined below.

Reviewer: Kenneth Fife

Minor Discretionary Comments:

1. For this study, the authors used data collected between 1999 and 2008. This is longer than any other previous NHANES/HSV 2 report. Presumably, the longer collection period was used to obtain enough individuals in the specific sub-categories for analysis. Although it is unclear what impact this may have on the current analysis, it may be worth pointing out this difference between this analysis and previous NHANES reports (probably in the discussion rather than the methods).

Response: The data used in our paper span a longer time period than other NHANES-based reports, as we are using newly released data from the most recent wave of data collection (2007-2008). For the most part, other researchers have not yet had a chance to publish manuscripts based on the new wave of data. In combining data collected over a 10-year period to obtain a large sample size, we are implicitly assuming that there have been no significant changes in prevalences of the variables measured, and associations between the variables of interest with HSV-2 seroprevalence. We have added this as a limitation in our discussion section.

2. The discussion is rather long and somewhat unfocused in places. Although length is not a particular consideration for this journal, it does obscure the more important points of the paper. For example, the discussion of the possible origins of the racial discrepancy in HSV 2 positivity rates, while interesting, contributes little to the discussion and the results presented in this manuscript contribute little new information on this topic.

Response: While results showing racial disparities in HSV-2 seropositivity may not be new information to researchers in the field, not every reader is familiar with this literature. Thus, we decided to retain the discussion, as it is important to provide some background and context related to our findings. We are concerned that if at least some discussion as to the differences between Blacks and other Americans is not provided, then our findings could be misused or misinterpreted. However, we did shorten this aspect of the discussion. The following sentences were thus deleted: “Of the almost 2.5 million people in state or federal prisons and local jails in 2008, almost 900,000 (36%) were Black males.[30] Higher mortality and
incarceration rates among young Black males contribute to low sex ratios, and higher rates of multiple partnerships among Black men than women.”

3. Conclusion (also the conclusion of the abstract) – The reference to the “viral nature of HSV 2” is a little puzzling. It is less the fact that genital herpes is caused by a virus than it is caused by a pathogen that causes chronic/recurrent infection that is novel. The “viral nature” of influenza A does not lead to similar problems.

Response: We acknowledge the chronicity of HSV-2 in the abstract and conclusion by adding “chronic” to the “viral nature of HSV 2”.

4. Table 1 – using percentages alone makes the table somewhat difficult to interpret. Is the fact that Black women between 40 and 44 have a lower percentage of HSV 2 positivity due to small numbers in that category? It cannot be determined. Consider adding some numbers (at least to, for example, the age categories) to the table.

Response: Numbers (i.e. row totals) were added to the age categories for Tables 1 and 2. This could not be done for Table 3 as the age categories are used to describe two different variables. Also, column totals were added to the relationship status categories for Tables 1 to 3.

5. Figure 1 – consider extending the y-axis to 100%.

Response: Done.

Reviewer: Amalia Magaret

Discretionary Revisions:

1. Discussion page 14: Regarding the last paragraph, it is also possible that the discrepancy between self-report and seroprevalence is independent of seroprevalence because the great majority of those infected may not recognize any potential symptoms as herpetic, though they can be instructed to do so. See Wald 2004 Herpes “Herpes Simplex Virus Type 2 and Transmission: Risk Factors and Viral Shedding” and Wald 2000 NEJM “Reactivation of genital herpes simples virus type 2 infection in asymptomatic seropositive persons”. Self-reported infection may simply indicate access to health care.

Response: Thank you for suggesting these articles to review for the Discussion section. The articles were read and one was cited within the mentioned paragraph.
2. Discussion page 15: Regarding the paragraph on taking herpes infection seriously, recent research has shown in several groups that the psychological impact of a new herpes diagnosis is mild and short-lived. See Richards 2007 STI “HSV-2 serologic testing in an HMO population” and (somewhat less relevant) Meyer 2005 “The psychosocial impact of serological herpes simplex type 2 testing in an urban HIV clinic”. The statement about “great psychological distress” should be qualified accordingly. Testing is an important aspect of prevention.

Response: Thank you again for suggesting these articles to review for the Discussion section. The articles were read and then cited within the mentioned paragraph.

3. Figure 1 would be clearer with four colors instead of 2. As it is, the comparison between seroprevalence and self-report is not easily done visually.

Response: Done.

Minor non-essential comments:

1. Move last sentence before the section “measures” to the beginning of the results section.

Response: Done

2. In the section “measures” a result is present: “In the course of our analysis, and consistent with published literature, we found differences in HSV-2 seropositivity and in the unadjusted rate of increase of prevalence with age between Black respondents and either Hispanic or White respondents.” This should be moved to “results”.

Response: This statement is present in this section due to the fact that it justifies the classification of our measures for the analysis in this paper. It was a finding that impacted our analysis, but not part of the final results presented, and its inclusion is necessary to understand the reasoning behind the progression of our analyses. Thus, we have retained it in the measures section.

3. Page 10, I think “For models within both age strata” means “For models within either age strata”. If this is correct, please modify as it would be clearer that a combined model was not performed.

Response: Done.

4. Discussion page 12: This sentence is particularly clear and helpful for interpretation of the findings: “We caution against the assumption of lower risk for a partner who is married or cohabitating … ”
Response: Thank you.

Reviewer: Maya Sternberg

Major Compulsory Revisions:

1. 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.

Response: The question we are addressing is predictive rather than etiologic. Specifically, we are interested in predicting prevalent infection and not HSV-2 incidence. In the fifth paragraph of the background section we explain that we hypothesize that prevalence may be higher in the pool of people from which new partners are most often drawn, those who are not currently married or partnered. This has relevance for new infection, in that prevalence in the partner pool plays a significant role in driving new infections within a demographic group. To clarify this further, we have made a number of small edits throughout, for example changing that statement that “Our findings indicate that relationship status provides important information about the risk of contracting HSV-2” to “Our findings indicate that relationship status provides important information about the risk of having contracted HSV-2.”

2. 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.
Response: Total lifetime partner number, as originally coded, was indeed not significantly associated with HSV-2 in our models. The fact that a variable is associated with both an exposure (relationship status) and an
outcome (HSV-2) does not ensure that it will act as an operational confounder in a model containing other covariates (e.g. age, sex, race). Thank you for drawing our attention to the non-linearity of the association between lifetime partner number and log odds of HSV-2. We have revised the analysis after recoding lifetime partner number into categories to avoid the linearity assumption. This allows us to generate a series of odds ratios that are more readily interpretable than a single OR for a log transformed variable. Your suspicions are correct. Lifetime partner number is now significantly associated with HSV-2 in both models, and this change is reflected in Table 4.

3. 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 ±1.96 (1.19) = (0.78, 83.0)!

With a confidence interval like this you may as well say you have no idea—whether it is large or not—

Response: We have not actually included any variables in our current model based on this criterion, and have deleted this statement. All variables included are highly significant (p<0.0001), including the omnibus test for the partner number categories.

4. 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).

Response: We have strengthened our argument, and revised this sentence. While the non-linear relationship – particularly what appeared to be a flattening of prevalence after age 30 – was what initially aroused curiosity, it was not the sole basis for the decision to stratify analysis on age. Since we were interested in relationship status, as part of our original model building we examined interactions between relationship status and the other demographic variables. The interaction between age and relationship status was significant (p<0.0001), as was the interaction between sex and relationship status (p=0.0071). Given our initial interest in age, we decided to stratify on that factor and continue to include an interaction term for sex and relationship status in the two age-specific models. The decision to stratify at a cut-point of age 30 was made based on 1) the shape of the relationship between age and HSV-2; 2) feasibility (in
that a group couldn’t be too small, and; 3) social (but not biological) differences between those in their 20’s versus those in their 30’s and 40’s.
5. 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 \( \exp(2.0892 - 0.0839(20)) = 1.3 \), 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.

**Response:** Thank you for catching this miscalculation. Given the change in our models with incorporation of a categorically coded lifetime partner number variable, this interaction is no longer in our model. However, we do have an interaction between age and race among those age 30 to 49, and have calculated the associated odds ratios correctly.

6. 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.

**Response:** These comments have been addressed above.

**Minor Essential Revisions:**

1. 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).

**Response:** We used the Wald chi-square test for our logistic regression analysis, and have added this detail to our manuscript.

Department of Epidemiology & Biostatistics • Schulich School of Medicine & Dentistry

The University of Western Ontario

Kresge Building • Room K201 • London, Ontario • N6A 5C1 • Canada

Telephone: (519) 661-2162 • Fax: (519) 661-3766 • www.schulich.uwo.ca/epidem
2. 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?

Response: Our original concern in conducting this analysis was to examine whether the epidemiologic pattern of HSV-2 seroprevalence levelling beyond young adulthood was a homogeneous pattern across relationship statuses. The concern here was that if those engaging in new partnerships had higher prevalence, this may have an impact on ongoing risk of transmission. As such, the group of never married, divorced, separated and widowed individuals represents the group that is most likely to be forming new sexual partnerships and is of equal interest.

Discretionary Revisions:

1. 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

Response: As per the comments of the other reviewers, we have revised our graphs. We have retained all the data in two tables, one for men and one for women, but used four rather than two colours and four distinctly shaped markers. Much of the cluttered look results from the overlap between the four lines of self-reported genital herpes frequencies, and this unfortunately remains in separated graphs unless we change the scale of the y axis, which then makes them less visually comparable with the HSV-2 seropositivity prevalences. We hope our revisions make the graphs as clear as possible, given the overlap.

2. 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

Response: Yes, we derived a set of 10-year weights as per NHANES guidelines. We have added a statement to our Discussion section to acknowledge the assumption that underlies combining across data waves regarding consistency of results across that time period.

Sincerely,
Greta Bauer, PhD, MPH
Nooshin Khobzi, PhD (cand)
Todd Coleman, PhD (cand)