

## Author's response to reviews

**Title:** Incidence of cytomegalovirus infection among the general population and pregnant women in the United States

### Authors:

Fernando AB Colugnati ([fcolugnati@uol.com.br](mailto:fcolugnati@uol.com.br))

Stephanie AS Staras ([SAS@ehpr.ufl.edu](mailto:SAS@ehpr.ufl.edu))

Sheila C Dollard ([sgd5@cdc.gov](mailto:sgd5@cdc.gov))

Michael J Cannon ([mcannon@cdc.gov](mailto:mcannon@cdc.gov))

**Version:** 3 **Date:** 16 May 2007

### Author's response to reviews:

BMC Infectious Diseases

Reply to Reviewers

Reviewer 1

"I think the symbols given in the legend of figure 1 are not those which are represented in this figure!"

We corrected the symbols in the legend so that they match.

Reviewer 2

"Although rightly stressing...the need of other than vaccine preventive measures for congenital CMV disease, the authors do not mention recent original or review papers on this matter, concerning protective behaviors and administration of CMV hyperimmunoglobulins or antiviral drugs for prevention or therapy of fetal infection and disease."

In the Discussion we added the references suggested as well as the sentence: "Such screening may also lead to the administration of CMV hyperimmunoglobulins or antiviral drugs for prevention or therapy of fetal infection and disease."

"Abstract and background: CMV is a common opportunistic pathogen (not infection)."

When talking about AIDS-related conditions, it is conventional to use the expression "opportunistic infection", so we left that phrase in the Abstract and Background.

"Symptoms may be absent during primary infection (see Nigro et...)."

Our sentence states that symptoms are usually absent. We feel that the literature is insufficient to make such a fine distinction, so we left the sentence as is. In addition, we did not add the suggested citation because we already had a citation supporting our statement.

"Pregnant women are at higher risk of acquiring infections because of estrogen/progesterone-induced immune depression."

We added the following phrase in the Discussion: "...and they may have a higher risk of acquiring infections because of pregnancy-induced immune depression." We also added one of the recommended citations.

"Figure 1: there is no correspondence in the symbols between legend and figure."

We corrected the symbols in the legend so that they match.

Reviewer 3

"NHANES III is well known, but the authors must provide the number of samples which were tested for CMV"

We included the number of specimens tested (N=11,859 for ages 12-49 and N=2,679 for ages 6-11) in the Methods section under the subheading Study population and design.

"There were large confidence intervals around the estimates of force of infection presented. Indicate which of the differences discussed in the text were statistically significant."

In part, the large confidence intervals were a result of the complex survey design. NHANES is a three stage stratified-clustered sample, where primary sample units are clusters generally defined by ZIP codes. So, despite the great number of subjects surveyed, they were not the units used in the statistical estimation.

This study design impacts the statistical methods used for any kind of estimation, either simple prevalence estimates or log-linear model parameters. The main impact is in the estimation of the variance-covariance structure. In classical statistics, where the sample units are assumed to be independently and identically distributed, we have  $N-1$  degrees of freedom, where  $N$  is the sample size. For complex surveys like NHANES, the  $N$  subjects are neither independently nor identically distributed. Degrees of freedom (df) vary according to the cluster and strata involved in the population subgroup of interest; roughly speaking the df is # of clusters - # of strata. Compared to a simple survey, variances are inflated, leading to higher standard errors and wider confidence intervals. In our study, variance inflation ranged from 2 to 8, depending on the population subgroup.

Furthermore, when we allowed the force of infection to vary for the younger age groups (6-11 years and 12-19 years), we were necessarily reducing the number of sample units from which force of infection estimates could be made. This also was a cause of higher variances for these age intervals and therefore wider confidence intervals.

We considered differences to be statistically significant when the corresponding confidence intervals did not overlap. We added the following sentence to the Methods section under the subheading Description of models: "We considered parameter differences to be statistically significant when corresponding confidence intervals did not overlap." We also added this sentence as a footnote to Table 1 and Figure 2.

"In Figure 1, the symbols on the legend to the Figure do not match those on the Figure itself."

We corrected the symbols in the legend so that they match.

"In Figure 2, it is surprising to see negative values for some estimates of force of infection given the presumed, large study size of the population (see above). Were these negative values still present when data sets were analysed without the piecewise log linear models introduced to control for apparent changes in slope of age-specific prevalence?"

We already discussed the issue of negative estimates for force of infection in the legend for Figure 2. A more detailed description and response follows.

The negative force of infection estimates occurred because we were treating a cross-sectional study as if it were a cohort study, in which seroprevalence is always monotonically increasing. Since that assumption was sometimes violated within certain subpopulations and narrow age intervals, it was possible to get negative forces of infection.

The negative values were not present when piecewise models were not used (i.e., when the model did not allow for different slopes within different age ranges). However, despite some force of infection estimates being negative, they were not statistically different from zero, which would be the best interpretation of these negative results.

To clarify this phenomenon, we added the following sentence to the Discussion: "Furthermore, the models implicitly assumed that seroprevalence was monotonically increasing with age, as if this cross-sectional study were a cohort study in which seroprevalence was measured at various ages of follow-up. However, this assumption was violated for some of the younger subpopulations. As a result, we occasionally obtained negative estimates for the force of infection (Figure 2), although these estimates were not statistically different from zero."

"In Table 2, the column headed "percentage seropositive" is redundant and should be deleted."

We deleted the column as requested.