1-20-2016

Models for HSV shedding must account for two levels of overdispersion

Amalia Magaret

University of Washington - Seattle Campus, amag@uw.edu

Models for HSV shedding must account for two levels of overdispersion

Amalia Magaret, PhD
Department of Laboratory Medicine , University of Washington

Contact:
Virology Research Clinic
Harborview Medical Center, Box 359928
325 9th Ave, Seattle, WA 98104
Phone: 206-520-4316
Fax: 206-520-4371

Abstract:

We have frequently implemented crossover studies to evaluate new therapeutic interventions for genital herpes simplex virus infection. The outcome measured to assess the efficacy of interventions on herpes disease severity is the viral shedding rate, defined as the frequency of detection of HSV on the genital skin and mucosa. We performed a simulation study to ascertain whether our standard model, which we have used previously, was appropriately considering all the necessary features of the shedding data to provide correct inference. We simulated shedding data under our standard, validated assumptions and assessed the ability of 5 different models to reproduce the parameters used and assess model performance. Our standard Poisson model, which includes a random per-person intercept to account for overdispersion, provided surprising findings: excess type I error (up to 28%) and poor coverage (~70%). A Poisson model with data-derived empirical variance structure, however, provided optimal power with type I error controlled at 5% and near 95% coverage. Explanations include that HSV detection frequency is doubly overdispersed: 1) person-level differences in shedding frequency must be accounted for by a random intercept and 2) the episodic nature of shedding induces additional extra-Poisson variance in person-level shedding frequency that must be accounted for using an empirical variance structure. These data support using the model with empirical variance structure for future crossover studies of HSV shedding.

Background

Herpes simplex virus (HSV) causes a lifelong infection. Following initial infection of mucosal epithelium, the virus resides in the neurons and frequently reactivates, emerging again at mucosal surfaces and causing ulcerations or lesions. The infection can be caused by either herpes simplex type 1 (HSV1, occurring in the mouth or genital region) or herpes simplex type 2 virus (HSV2, commonly occurring genitally). As many as 11% of the world's population may be infected with HSV2, and 67% with HSV1[1,2]. When virus is detected on the skin or mucosa, it is often associated with symptoms or signs; but the virus is also frequently present in the absence of oral or genital lesions.

Genital viral shedding is an efficient and powerful outcome for assessing herpes disease severity, and an accurate surrogate marker in studies evaluating the efficacy of candidate antiviral agents[3]. HSV shedding is more common than lesions, can be assessed by a laboratory measure that is more accurate than patient or clinician assessment of lesions, and correlates with frequency of genital lesions. Importantly, the strong within-person consistency makes it a meaningful and reliable measure upon which to base evaluation of disease severity. As such, we have conducted several studies of interventions designed to demonstrate benefit for HSV infections by using frequency of HSV detection (viral shedding) as the primary outcome[3-6]. In these studies participants obtain daily swabs from the genital mucosa which are tested for HSV by PCR, and the information is summarized as the proportion of swabs with detectable HSV DNA. Observed reductions in frequency of shedding induced by interventions provide information regarding efficacy that can be further confirmed through studying clinical outcomes such as frequency of genital lesions in larger Phase III studies.

There are two sources of variability in HSV mucosal shedding. The first is wide variation in shedding rate noted between persons that reflects the unique host-pathogen interaction within the individual[7,8]. The second variability results from the occurrence of shedding in clusters of days, followed by shedding-free intervals[9]. Therefore, the risk of shedding on a given day depends not only on the shedding rate inherent for a person, but also on whether shedding occurred the prior day.

In the current simulation study, we evaluate a range of analytic methods to address viral shedding as an outcome in cross-over clinical trials. Our findings result in a recommendation for an optimal method to be used in currently ongoing and forthcoming studies.

Methods

Design choice: the crossover

Crossover designs are efficient in making use of the demonstrated consistency in person-specific HSV detection frequencies. When comparing individuals to themselves in a crossover design, the variability of the effect size is reduced, relative to cohort designs that make comparisons between groups. Consequently, we have made use of the efficiency of within-persons comparisons by conducting crossovers, in which all persons receive, sequentially, the intervention followed by or preceded by a standard comparator (usually placebo). Shedding frequency is subsequently compared between the two treatment periods, and the difference is described as a ratio of

2

shedding rates. When an investigational treatment is compared as such to a standard treatment, the shedding rate ratio can be less than 1, describing a proportionate reduction in shedding.

<u>Model choices</u>
In order to estimate average rates of shedding on both arms and to estimate the shedding rate ratio, we have utilized Poisson regression. Poisson regression assumes that the variability of responses is equal to the average response[10]. In HSV2 infections, person-to-person differences in shedding rates indicate that there is no single average response rate, but that each individual responds to HSV infection with a unique severity level. Accordingly, we have historically accounted for this variation by including in regression models an adjustment that allows each person to have their own baseline shedding rate estimate, against which the shedding rate while on treatment is compared. This was done using a random intercept. This intercept removed the assumption of constant variance as well. Remaining error variance was then estimated using model-based methods, assuming the variance could be estimated using the within-person response rate.

In this simulation study, we compare our standard model against 4 other models, listed here and further detailed below. These models represent the most likely options that take into account all the features of the data. Not included are marginal models, such as generalized estimating equations, that do not make use of the interest in assessing within-person differences by arm but instead compare average outcomes on each arm. Also not included are linear regressions that would assume shedding rate can take all possible real numbered values.

- **Poisson** regression with random intercept and **empirical** variance structure
- **Negative binomial** regression (allows extra Poisson variance per individual)
- **Negative binomial** with random intercept and **empirical** variance structure
- **Baseline covariate** model that includes baseline shedding as predictor

All four of these new models relax the assumption that, even at the individual level, the variance could be estimated by the average shedding rate. Empirically estimated variance allows error variance to be estimated from the observed data itself, negative binomial models also utilize observed data and allow for variation inflation at the individual level, assuming the variance is proportional to (but may be larger than) the average outcome rate.

All but the *baseline covariate* model are random effect models, so they allow two rows of data per participant in the study: one row per person for each arm. They can generally be designated using the following notation, where Y_{ij} is the number of positive swabs for person i on arm j, n_{ij} is the number of swabs taken for person i on arm j, T is an indicator for arm (1=investigational treatment of interest, 0=comparator or placebo), γ_i is the random person-level intercept and ε_{ij} is the error term. The computation of both $\gamma_i + \varepsilon_{ij}$ differ depending on the specification of the variance structures as described above, however the standard model and the first three proposed potential models share this structure:

$$\log(Y_{ij}) = \alpha + \beta*T + \log(n_{ij}) + \gamma_i + \varepsilon_{ij}$$

We can expand to show that the log shedding rates on the two treatment arms differ by the coefficient β and random error.

$$\log(Y_{i0}/n_{i0}) = \alpha + \gamma_i + \varepsilon_{i0}$$

3

$$\log(Y_{i1}/n_{i1}) = \alpha + \beta + \gamma_i + \varepsilon_{i1}$$

Rearranging then and acknowledging that the errors have mean zero, we can see that β is the estimate of the log shedding rate ratio, or equivalently that the shedding rate ratio is estimated by e^β.

$$E[\log(Y_{i1}/n_{i1}) - \log(Y_{i0}/n_{i0})] = E[\log(Y_{i1}/n_{i1} \div Y_{i0}/n_{i0})] = \beta$$

The final *baseline covariate* model, however, includes only one row per participant, and includes offset terms for both the log shedding rate during the comparator arm and the log number of swabs while on the investigational arm. Here we estimate the variance ε_i by allowing overdispersion and including the standard residual-derived scale parameter, again not relying on the distributional assumptions of a Poisson model.

$$\log(Y_{i1}) = \alpha + \log(Y_{i0}/n_{i0}) + \log(n_{i1}) + \varepsilon_i$$

Rearranging the above expression we see that the interpretation of α is of the estimated shedding rate ratio, intervention versus run-in, and therefore is has the identical interpretation to that of the β term in the previous models.

$$\log(Y_{i1}/n_{i1}) - \log(Y_{i0}/n_{i0}) = \log(Y_{i1}/n_{i1} \div Y_{i0}/n_{i0}) = \alpha + \varepsilon_i$$
$$E[\log(Y_{i1}/n_{i1} \div Y_{i0}/n_{i0})] = \alpha$$

The advantage of the *baseline covariate* model is that fewer assumptions are required: there is no person-to-person correlation when each person contributes only one data point. The disadvantage is that the offset $\log(Y_{i0}/n_{i0})$ is often undefined in real data ($\log(0)=-\infty$), as perhaps 30% of participants do not shed over the usual sampling period of about a month and so for them $Y_{i0}=0$. For our purposes in simulation, we reassigned the value of Y_{i0} to be 0.5 whenever the observed Y_{i0} was 0. This allowed us to compute a valid $\log(Y_{i0}/n_{i0})$, but likely biased findings by inflating the observed shedding rate while not on intervention.

Simulation specifics

We simulated and tested each condition 1600 times to allow approximately 1% precision for our estimates of power and type I error. For each, we imposed the following characteristics on the data. The parameters describing shedding characteristics were previously shown to correspond well to clinical data from a large cohort of HSV2 seropositive persons[11,12].

- 20 - 50 persons enrolled in a crossover study
- 28 days of samples collected for HSV detection, per person on each of two arms
- A 13% average baseline shedding rate, allowing individual variation in true rate using a β prior with standard deviation of 12% ($\alpha = 1.0$, $\beta = 6.2$),
- Auto-regressive correlation to create shedding episodes, with $\phi = 0.54$ the average correlation in the probability of shedding between two consecutive days for the same person. ϕ also employed a beta prior, standard deviation 0.23, ($\alpha = 2.1$, $\beta = 1.8$).
- An intervention with 0%, 20%, 40%, 60% or 80% efficacy (corresponding to shedding rate ratios of 1.0, 0.8, 0.6, 0.4, and 0.2, respectively).

Definitions

We define power as the probability of detecting a statistically significant effect of the intervention when in fact that intervention is effective (has an imposed shedding rate ratio not equal to 1.0). We define type I error as the probability of detecting a statistically significant effect of the intervention when the intervention is *not* effective (an imposed shedding rate ratio of 1.0).

4

Coverage is the probability that the model-derived 95% confidence interval includes the true parameter: for our purposes we tested coverage of the intervention efficacy. Monte-Carlo confidence intervals for the intervention efficacy are computed using the standard error of the shedding rate ratio estimates over repeated simulation. Model-based confidence intervals do not contribute to this computation.

Results

Simulation results when 50 persons per arm were simulated are shown in Figures 1, 2 and 3. Figure 1 shows that both the Poisson (standard, in brown) method with the random intercept and the Poisson method that also includes the empirical variance estimate (purple) provide the most accurate estimates for treatment efficacy, as they are closest to zero bias.

Figure 1. Estimates and 95% confidence intervals for mean treatment effect; they are based on variance over the simulations (Monte-Carlo variation).
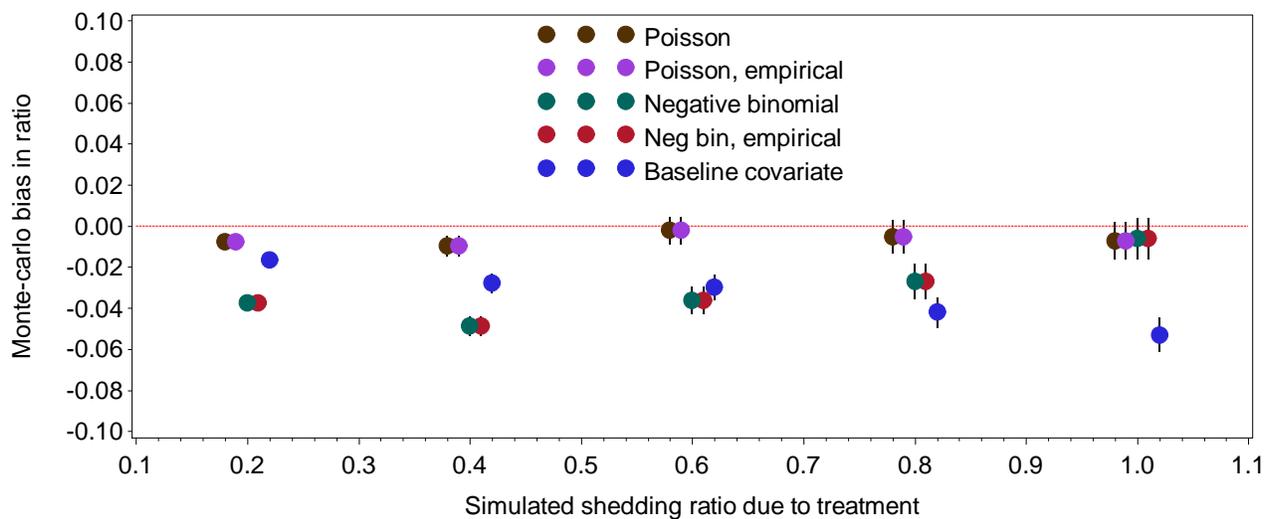

Figure 2 shows, in red lines, the targeted range for coverage, which is 95% plus and minus the expected Monte-Carlo error of 1.6% based on 1600 simulated datasets. Bars with heights that fall between the red dashed lines have coverage that would be consistent with the anticipated 95% coverage. While most analytic methods are close to 95% coverage, the standard method is closer to 70% coverage, indicating that about 30% of the time, the 95% confidence interval for the true intervention efficacy does not include the true value.

5

Figure 2. Coverage: The y-axis shows the proportion of times that the model-estimated confidence interval for the treatment effect *actually includes* the simulated treatment effect. We would expect 95% coverage from all models.
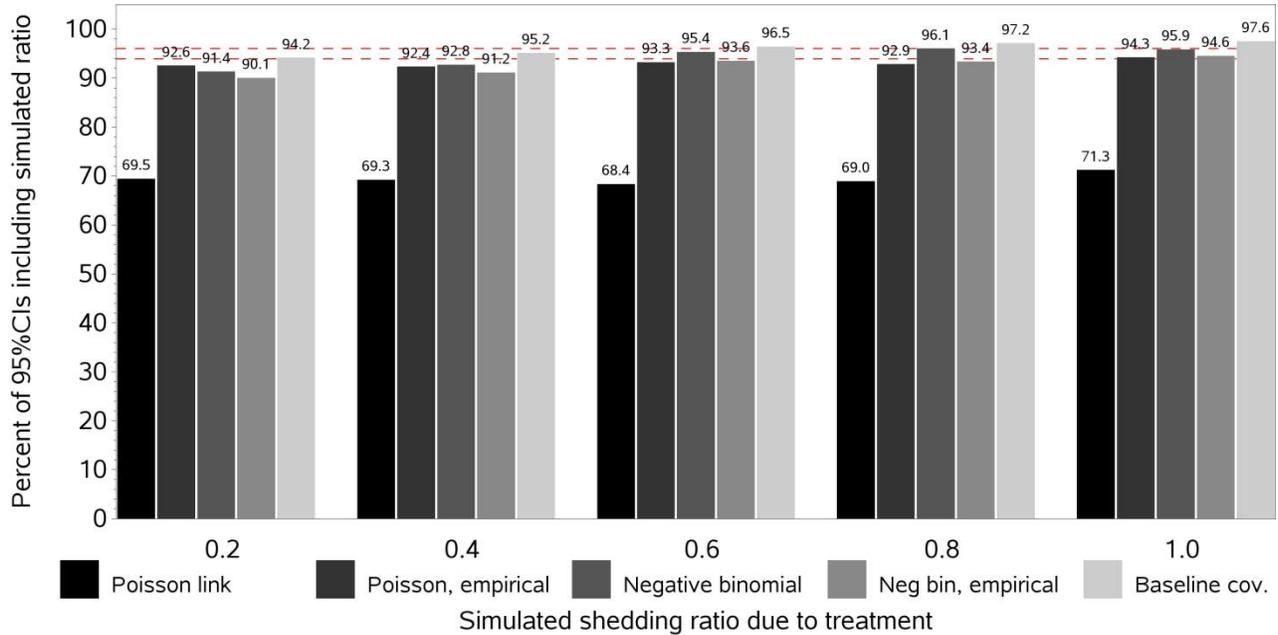

In Figure 3 we examine power or Type 1 error rate, and show that when the imposed or simulated shedding rate ratio is actually 1, indicating no benefit of treatment, the standard Poisson regression method with random intercept provides an almost 28% probability of detecting a statistically significant treatment effect. This type I error can be interpreted to mean that when the treatment has no impact on the outcome rate, there is still an approximate 1:4 chance that a significant p-value (<0.05) will result from the study.

Both analytic methods using the empirical variance structure (Poisson and negative binomial) provide type I error of approximately 5%. Since these methods control type I error, we refer back to Figure 1 showing Poisson and Poisson empirical to be most accurate of all methods tried, and conclude that Poisson regression with empirical variance structure performs ideally among methods tried. Looking back again to figure 3, and comparing power this time, the power for the Poisson regression with empirical variance estimate (second darkest bar) is comparable to other models tested, when considering those that control type I error at 5%. (Note that power can fall below the standard target of 80% for these simulations, but relative relationships between methods are consistent.)

Results were similar when 20 persons were used, with nearly identical rankings of the methods in terms of power, coverage and bias (data not shown).

6

Figure 3. Power, or type I error rate, depending on what is true. Each bar shows the proportion of times the p-value is less than 0.05, that is, a significant treatment effect is detected. For simulated shedding ratios less than 1.0, the bar height is interpreted as *power*, as there is a true treatment benefit. When the simulated ratio equals 1.0, the bar height is interpreted as *type I error*. The red lines show the anticipated range for type I error, which is 5% plus and minus the expected Monte-Carlo error of 1.6%. Bars whose height falls between the red dashed lines have a type I error that is consistent with 5%.
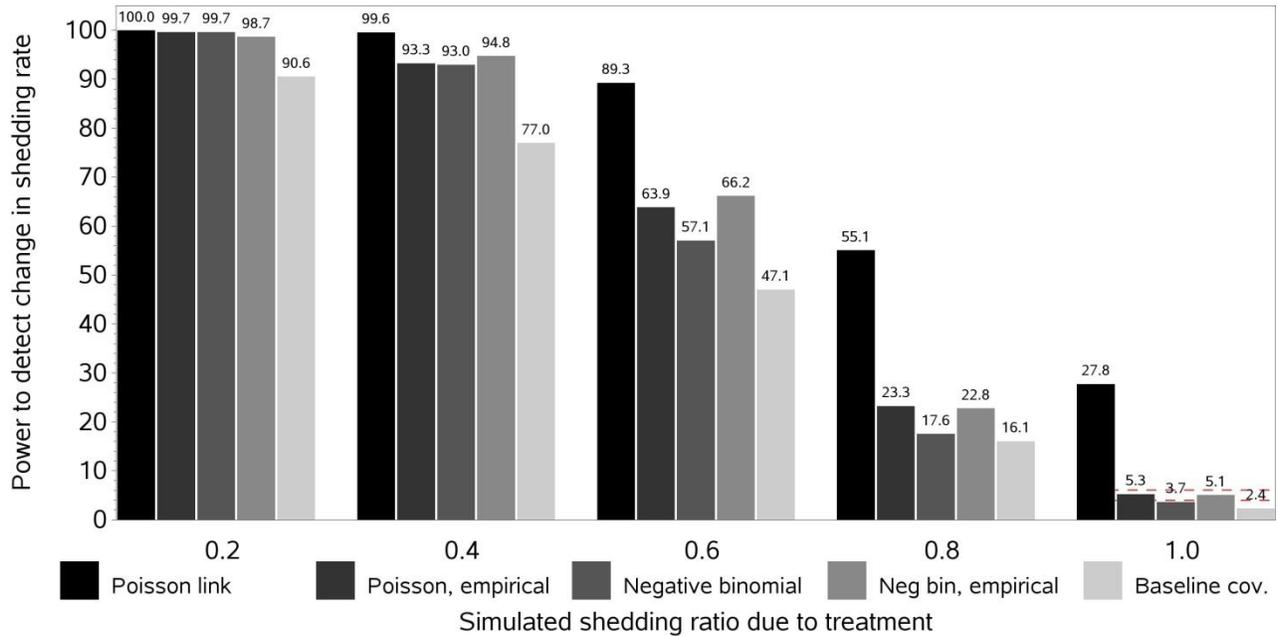
Conclusions

In this simulation study, we determined that shedding frequency data involves two nested levels of extra-Poisson variance (overdispersion), one at the population level and one at the individual level, and both needed to be accounted for in the analysis. While standard Poisson random effects model which we have used in our prior work accounted for the population level overdispersion, that model was not sufficient to account for the additional increased variability at the individual level due to the clustering of days with shedding (i.e, shedding episodes).

We have demonstrated for the first time that previously used regression techniques for analyzing shedding rate resulted in inflated type I error rates of as high as 28%. We were surprised to find that shedding episodes had such a dramatic effect on variance estimates that they could result in false inference. This type I error can be interpreted to mean that when the treatment has no impact on the shedding frequency, there is still an approximate 1:4 chance that a significant p-value of less than 0.05 will result from the study. This is different from concluding that a significant p-value has a 1:4 chance of being spurious (this is inverted conditional logic). The knowledge of an inflated type I error leads us to consider p-values obtained from this previous standard method with concern. Consequently, we are in the process of re-analyzing previously published studies.

Importantly, we note that the effect estimates obtained from the Poisson regression without empirical variance structures do not change when those data are subjected to the new model. However, both standard error estimates and the corresponding confidence intervals are likely to widen, with p-values being simultaneously likely to increase in value (and decrease in significance).

In the future, we plan to adopt this Poisson regression with a random intercept and empirical variance estimate, as an analytic approach with optimally desirable properties. This new method assures us that model-generated confidence intervals provide the stated level of confidence. We can be certain that p-values generated under the new model are only likely to be significant under the null model fewer than 5% of the time, the desired error rate. And we are also certain to have power about as high as can be obtained, since we have considered all reasonable candidate models likely to be implemented. The simulation results reassure us that we have identified a method with predictable and desirable properties to implement in future studies.

8

References

1.   Looker KJ, Magaret AS, May MT, et al. Global and Regional Estimates of Prevalent and Incident Herpes Simplex Virus Type 1 Infections in 2012. *PLoS One* 2015; **10**(10): e0140765.

2.   Looker KJ, Magaret AS, Turner KM, Vickerman P, Gottlieb SL, Newman LM. Global estimates of prevalent and incident herpes simplex virus type 2 infections in 2012. *PLoS One* 2015; **10**(1): e114989.

3.   Wald A, Corey L, Timmler B, et al. Helicase-primase inhibitor pritelivir for HSV-2 infection. *N Engl J Med* 2014; **370**(3): 201-10.

4.   Gupta R, Wald A, Krantz E, et al. Valacyclovir and acyclovir for suppression of shedding of herpes simplex virus in the genital tract. *J Infect Dis* 2004; **190**(8): 1374-81.

5.   Johnston C, Saracino M, Kuntz S, et al. Standard-dose and high-dose daily antiviral therapy for short episodes of genital HSV-2 reactivation: three randomised, open-label, cross-over trials. *Lancet* 2012; **379**(9816): 641-7.

6.   Mark KE, Corey L, Meng TC, et al. Topical resiquimod 0.01% gel decreases herpes simplex virus type 2 genital shedding: a randomized, controlled trial. *J Infect Dis* 2007; **195**(9): 1324-31.

7.   Phipps W, Saracino M, Magaret A, et al. Persistent genital herpes simplex virus-2 shedding years following the first clinical episode. *J Infect Dis* 2011; **203**(2): 180-7.

8.   Tronstein E, Johnston C, Huang ML, et al. Genital shedding of herpes simplex virus among symptomatic and asymptomatic persons with HSV-2 infection. *JAMA* 2011; **305**(14): 1441-9.

9.   Schiffer JT, Wald A, Selke S, Corey L, Magaret A. The kinetics of mucosal herpes simplex virus-2 infection in humans: evidence for rapid viral-host interactions. *J Infect Dis* 2011; **204**(4): 554-61.

10.   Diggle PJ, Heagerty PJ, Liang KY, Zeger SL. Analysis of longitudinal data. 2nd ed. Oxford: Oxford University Press; 2002.

11.   Magaret AS, Johnston C, Wald A. Use of the designation "shedder" in mucosal detection of herpes simplex virus DNA involving repeated sampling. *Sex Transm Infect* 2009; **85**(4): 270-5.

12.   Magaret A, Stanaway J. Sample Size for a Binomial Proportion with Autocorrelation. *Statistical Communications in Infectious Diseases* 2011; **3**(1): Article 8.